CONTENTS OF VOLUME 78

Whole No.

578 Information Processing in the Child: Significance of Analytic and Reflective Attitudes. Jerome Kagan, Bernice L. Rosman, Deborah Day, Joseph Albert, and William Phillips.

579 Idiodynamic Sets in Word Association. Louis J. Moran, Roy B. Mefferd, Jr., and James P. Kimble, Jr.

580 Experimental Studies of Associative Variables in Originality. Irving Maltzman, Marigold Belloni, and Martin Fishbein.

581 An Information-Processing Model of One Kind of Problem Solving. Edward S. Johnson.

582 Cue Probabilism and Inference Behavior. Lee Roy Beach.

583 Recognition, Assimilation, and Identification of Objects. Lee Roy Beach.

584 Relating Personality and Biographical Factors to Scientific Creativity. J. A. Chambers.

585 Conflict and Stress as Related to Physiological Activation and Sensory, Perceptual, and Cognitive Functioning. Walter D. Fenz.

586 Uncertainty, Structure, and Preference. Harry Munsinger and William Kessen.

587 Physiological Activity and Performance: A Comparison of Kindergarten Children with Young Adults. Rogers Elliott.

588 Psychotherapy Experience in Retrospect: Questionnaire Survey of Former Patients and Their Therapists. Hans H. Strupp, Martin S. Wallach, and Michael Wogan.

589 Visual Perception in Strabismus and Amblyopia. Samuel C. McLaughlin.

590 Improving the Academic Performance of Anxious College Freshmen: A Group-Counseling Approach to the Prevention of Underachievement. Charles D. Spielberger and Henry Weitz.

591 Sex Differences in Body Perception. Seymour Fisher.

592 Varieties of Projection in Trait Attribution. Donald T. Campbell, Norman Miller, Jacob Lubetsky, and Edward J. O'Connell.

593 Extinction of Associations in Four Transfer Paradigms. Jean Barnes McGovern.

Vol. 78, No. 1 Whole No. 578, 1964

Psychological Monographs: General and Applied

INFORMATION PROCESSING IN THE CHILD:

SIGNIFICANCE OF ANALYTIC AND REFLECTIVE ATTITUDES[1]

JEROME KAGAN, BERNICE L. ROSMAN,[2] DEBORAH DAY,
JOSEPH ALBERT,[3] AND WILLIAM PHILLIPS
Fels Research Institute, Yellow Springs, Ohio

A series of studies with children in Grades 1–4 inquired into the immediate and historical determinants of a preference for analytic conceptual groupings. An analytic concept is based upon shared similarity in a particular objective component among a set of stimuli (e.g., animals with 1 ear, people with hats on). Results revealed that 2 more fundamental cognitive dispositions each contributed variance to the production of analytic concepts: the tendency to reflect over alternative solutions or classifications in situations in which several response alternatives are available simultaneously, and the tendency to analyze visual arrays into their component parts. These 2 dispositions are relatively independent of each other, orthogonal to verbal skills; and each influences the frequency of errors in perceptual recognition tasks. Degree of reflection over alternative solution hypotheses (as measured by response time) displayed remarkable generality across a variety of tasks and marked intraindividual stability over a 1-yr. period. There was an inverse relation between the production of analytic concepts and extreme degrees of hyperactivity and distractibility contemporaneously as well as during the 1st 8 yr. of life.

PROBLEM solving involves at least three distinct processes that occur in the following sequence: the initial categorization of information, storage of the encoded information, and the imposing of transformations or mediational elaborations upon the encoded material. The form of the initial categorization, transformation, or mediation is governed, of course, by the nature of the problem. Students of cognitive development have generally assumed that the striking differences between the cognitive products of children of different ages (or among children of the same age) were due primarily to differences in the availability of a classification vocabulary, transformation rules, and mediational diversity. In effect, the superior problem solving performance of older, in contrast to younger, children has been assumed to result from the richer knowledge repertoire of the older children. This supposition is intuitively attractive and empirically verified. It is not surprising, therefore, that psychologists have not seriously entertained the possibility that other factors may contribute to age and individual differences in cognitive products. Specifically, there has been a tendency to neglect the relevance of individual differences in the processing of information—differences in the aspects of stimuli that are initially selected for labeling and the degree of reflection attendant upon classification of events or the selection of solution hypotheses. For children and adults have a clear preference hierarchy with respect to the stimulus characteristics to which they initially attend and the speed with which classification decisions are made.

The classical dimension of analytic versus global categorizations describes a stimulus processing variable. The tendency to analyze a stimulus into its differentiated components increases with age, and at any single age there are stable individual differences in this tendency (Wohlwill, 1960). The notion of a generalized disposition to analyze external experience is not a new one. Early investigators in psychophysics noted that "analytic" subjects produced more sensitive threshold values than "synthetic" subjects. More recently, Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954) and Witkin, Dyk, Faterson, Goodenough, and Karp (1962) have used the phrases *analytic* and *field independent* in somewhat similar ways and have demonstrated that the tendency to analyze complex stimulus situations and detect patterns embedded in a distracting context increases with age and predicts basic behavioral characteristics of the individual.

---

[1] This research was supported, in part, by Grant M-4464 from the National Institute of Mental Health, United States Public Health Service. The authors wish to express their gratitude to the officials of many schools who have made this research possible including H. Baker, H. Brown, G. Canaday, Wilma Chrisman, P. Coulter, H. Cross, Rachel Dennison, W. Hamilton, and H. Thornburg.

[2] Now at the Department of Psychiatry, School of Medicine, Yale University.

[3] Now at the University of Illinois.

A central problem in discussing this dimension is the existence of markedly different sets of measurement operations, each of which has been labeled as an index of an analytic tendency. These varied test procedures are probably not assessing identical constructs, and it is necessary to restrict statements about analytic tendencies to the operations used in any particular study.

Analytic Concepts: Background Information

A series of investigations with school age children revealed that some children showed a clear preference to group familiar objects together on the basis of objective elements of similarity that were component parts of the total stimulus (Kagan, Moss, & Sigel, 1963). The technique used to measure this preference for analytic conceptualizations was a set of 30 cards, each with three black and white drawings of familiar objects (hereafter called the Conceptual Style Test or CST). Figure 1 illustrates some stimuli from this instrument.

The child was asked to "Pick out two pictures that are alike or go together in some way" and to state the basis for his grouping. These stimuli were constructed to elicit the two major concept classes of analytic and relational concepts, and, to a lesser degree, inferential-categorical concepts. An analytic concept involved pairings based on similarity

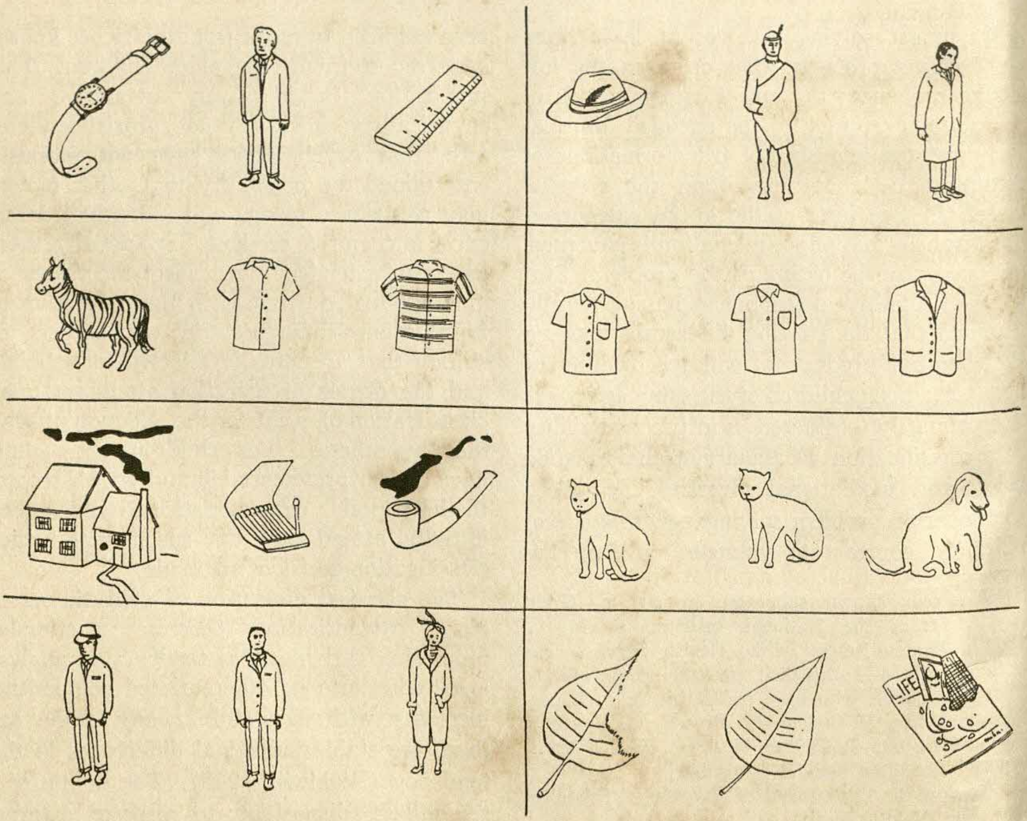


FIG. 1. Sample items from CST.

in an objective attribute that was a differentiated part of the total stimulus. Examples of this category are, "The watch and ruler have numbers," "The zebra and shirt have stripes," "The house and pipe have smoke coming out," "People with hats on," "Animals with their tongues out." The basis of similarity in each of these examples involves a component part of the total stimulus (e.g., numbers, stripes, smoke, hats, tongues). The relational category involved pairings based on a functional relationship between two stimuli. For these particular pictures, a functional relationship was an obvious association to the test item. Examples of relational concepts are, "The man wears the watch," "The matches light the pipe," "The hat goes on the man," "The dog and cat play together." The inferential-categorical class included conceptual pairings that were based on similarity in some inferred quality or involved a language convention. Examples of inferential concepts are "Inanimate objects," "Articles of clothing," "Human beings."

Earlier investigations indicated that the tendency to delay a conceptual decision was associated with the production of analytic concepts on CST (Kagan et al., 1963). The material in this monograph considerably extends the earlier studies. The data demonstrate the consistency of an analytic attitude across situations and suggest that the more fundamental processes of reflection versus impulsivity and visual analysis of complex arrays are primary determinants of the production of analytic concepts.

Developmental Changes in Analytic Concepts. CST has been administered to over 800 children in the elementary grades, and Figure 2 shows the linear increase in analytic responses with age. The data in Figure 2 are based on individual administration of CST with the exception of the Grade 4 data marked SMSG.[4]

First graders typically produced an average of about 4 analytic responses; sixth graders produced 9 to 10 analytic responses.

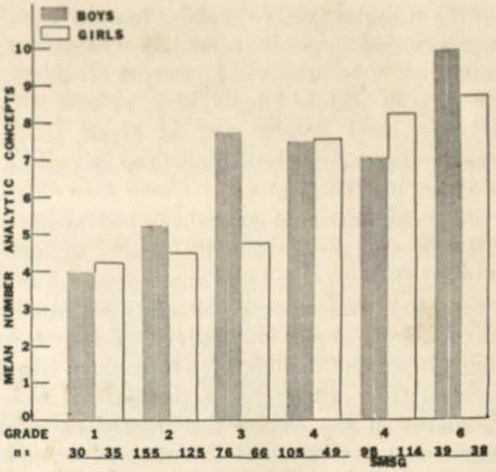


FIG. 2. Age changes in analytic concepts on CST.

Stability of an Analytic Attitude. The continuity of an analytic conceptual preference over a 12-month period was determined in three small samples. For one group of children tested in Grade 3 and then again in Grade 4, the coefficient was .70 for 24 girls and .43 for 22 boys. For another group of 35 boys tested first in Grade 4 and then in Grade 5, the coefficient was .73. Finally, for 52 boys originally tested in Grades 1 and 2 and retested in Grades 2 and 3, the stability coefficient was .47. There was, therefore, moderate stability over a 12-month period and increasing stability of an analytic attitude with age. The split-half reliability for this 30-item test was excellent. The corrected split-half reliability coefficient based on 300 protocols was .94.

Meaning of Analytic Concepts. Previous investigations (Kagan et al., 1963) indicated that children who were highly analytic on the 30-item CST, in comparison to nonanalytic children, were more likely to (*a*) produce interpretations of ink blots that were highly differentiated and contained references to descriptive aspects of the stimulus (e.g., butterfly with wings and feet) and produce fewer indistinct percepts (e.g., mud, cloud, paint), (*b*) attach a new verbal response to the component parts of a geometric design rather than only to the most salient aspect of the stimulus, and (*c*) inhibit impulsive and uncontrolled motor behavior. Moreover, analytic boys acquired concepts requiring analysis of picture stimuli (e.g., objects with missing legs; objects with a black band) earlier than nonanalytic boys (Lee, Kagan, & Rabson, 1963).

---

[4] Grade 4 SMSG data came from a group administered procedure using a booklet of 16 pictures, and administered as part of the National Longitudinal Study of Mathematical Abilities, sponsored by the School Mathematics Study Group, directed by E. G. Begle of Stanford University.

The demonstrated generality of an analytic attitude was an incentive for further exploration of the correlates and antecedents of this conceptual preference, particularly the ability to reflect over alternative classification hypotheses and the tendency to analyze visual arrays. The studies to be summarized were attempts at uncovering the antecedents and significant correlates of this domain of conceptual behavior.

STUDY 1

This study assessed the relation between analytic concepts on CST and performance in three situations requiring analysis of complex visual patterns: (*a*) the ability to locate meaningful stimuli that were embedded in a context of interfering lines, (*b*) performance on special subtests of the Wechsler Intelligence Scale for Children (WISC), and (*c*) degree of articulation present in interpretations of ink blot stimuli. The second purpose was to attempt to replicate an earlier finding that analytic children were more likely than nonanalytic ones to respond to nouns and adjectives in a word association test with words that belonged to the same syntactic class as the stimulus. This tendency was not highly correlated with IQ or vocabulary score. One possible rationale for this result was that analytic children might be precocious in learning syntactic classes and in learning that nouns have something in common with other nouns. It is not completely fanciful to speculate (*a*) that the learning of syntactic equivalences involves an implicit analysis of speech, noticing that nouns appear first in sentences and are given vocal emphasis and (*b*) that adjectives like "hot" and "black" are often contrasted with their opposites "cold" and "white."

Procedure

Subjects. The subjects (*S*s) were 62 boys (32 in Grade 1 and 30 in Grade 2), Caucasian, middle class, and of average intelligence. The age range for Grade 1 was 6-1 to 7-4 with a median of 6-9; for Grade 2 the range was 7-4 to 8-7 with a median of 7-9. Each *S* was seen on two separate occasions at the Institute by a female *E* (BLR), and 56 of the 62 subjects were seen for a third session with a male and female *E* present (BLR and JK).

Administration of Tests: Session I.

1. CST—The *E* said,

> I have some cards here with three pictures on them like this one. I want you to pick out two pictures that are alike or go together in some way. [After the child had selected two pictures *E* asked him] "Why did you pick these"?

The variable of primary interest was the number of analytic responses (i.e., a conceptual pairing based upon similarity in component parts).

2. Word-association test—In this test 40 words (20 singular nouns, 10 adjectives, and 10 verbs) were presented to each *S* with the following instructions.

> I am going to say a word out loud and you tell me the first word that comes into your mind as quickly as you can.

Response latency and the child's response were recorded. The child's meaningful associations fell into several major classes. Most of the responses to the nouns were (*a*) nouns in the same conceptual class as the stimulus, (*b*) verbs that were phrase completions to the nouns, or (*c*) nouns that were functionally related to the stimulus noun. Noun responses in the same conceptual class (called coordinate nouns) were either synonyms of the stimulus or words that shared membership in a conceptual class with the stimulus. There were, of course, noun responses that were neither synonyms nor instances of the same class (i.e., dog-fire, glass-car). The response CAT to the stimulus DOG was scored as a coordinate noun for both were members of the class ANIMAL or PET. The response MILK to GLASS was not scored as a coordinate noun because glass and milk have a functional relationship and are not independent instances of a superordinate category.

The response of major interest to adjectives was production of antonyms (e.g., dirty-clean, tall-small, hot-cold). The other categories that occurred were either nouns or synonyms. The responses to verbs were either verbs or nouns. The variable of major interest was a verb response that belonged to the same conceptual class as the stimulus verb (called *coordinate verbs*, e.g., throw-catch, run-walk, run-trot, write-draw, draw-paint, jump-stand, walk-sit).

The second grade boys produced significantly more coordinate noun, coordinate verb, and antonym responses than the first graders. The mean scores for the three categories were 8.2 versus 3.4, 3.6 versus 1.0, and 6.2 versus 1.8 ($p < .01$ for each comparison). Unless stated otherwise, all p values in this monograph are for two-tailed tests, and significance levels for mean differences are based on t tests. The year of development between Grades 1 and 2 is witness to a large increase in the tendency to produce word associates that match the syntactic class of a stimulus. Additional data on the frequency of coordinate nouns, verbs, and antonyms for 107 third and 52 fourth grade boys indicated little change in frequency of occurrence of these categories between these two ages. The means for the third and fourth grade *Ss* were 9.8 versus 9.5 for coordinate nouns, 5.1 versus 5.6 for coordinate verbs, and 5.6 versus 5.4 for antonyms. The mean scores for Grades 3 and 4 were similar in magnitude and slightly higher (for nouns and verbs) than the scores for Grade 2 children. There is, therefore, a more significant shift in the syntactic character of word-association responses between Grades 1 and 2 than between Grades 2–3 or 3–4.

3. Hidden Figures Test (HFT)—HFT was devised especially for this study and consisted of one practice and 11 test items. Each item included a figure card that illustrated a familiar object (e.g., Mickey Mouse, a cowboy, a rose), and a hidden figure card in which the familiar object was embedded (i.e., blended into a patterned background). Sample items appear in Figure 3.

Administration. The *E* said, "Here is a picture. What is this?" On the rare occasion when the child did not respond correctly *E* said, "It is a——." Then the hidden figure card was placed in front of the child to the right of the figure card and *S* was told, "Now you find the —— in this picture. Point to the —— when you find it."

If *S* selected incorrectly, he was told that he had not found the figure and was encouraged to look for it again. A maximum of 60 seconds was allowed for each item. The number of incorrect guesses, response time to the correct selection, and success or failure on each item were recorded. The items differed markedly in difficulty, and a weighted score for each child's performance was computed using weighted solution times for each item. The odd-even intratest reliability coefficients, corrected by the Spearman-Brown formula, were .78 for the first grade and .41 for the second grade. A frequency distribution of solution times was constructed for each item for each grade, and each of the 11 distributions was divided into thirds. A score of 3 was given for the shortest solution time, 2 for the moderate solution times, and 1 for the longest solution times and failures. Each child was given a score of 1, 2, or 3 for each item, and *S*'s total score was the sum of the 11 item scores. The median solution times for the 11 items prior to weighting revealed that the second grade children solved these hidden figures significantly earlier than the first grade children (2.4 seconds versus 4.7 seconds; $p < .001$).

A second variable obtained from HFT was the number of incorrect solutions offered by the child. The child was allowed to offer solutions until he solved the item (or for a maximum of 60 seconds) and the range of scores on this variable was considerable.

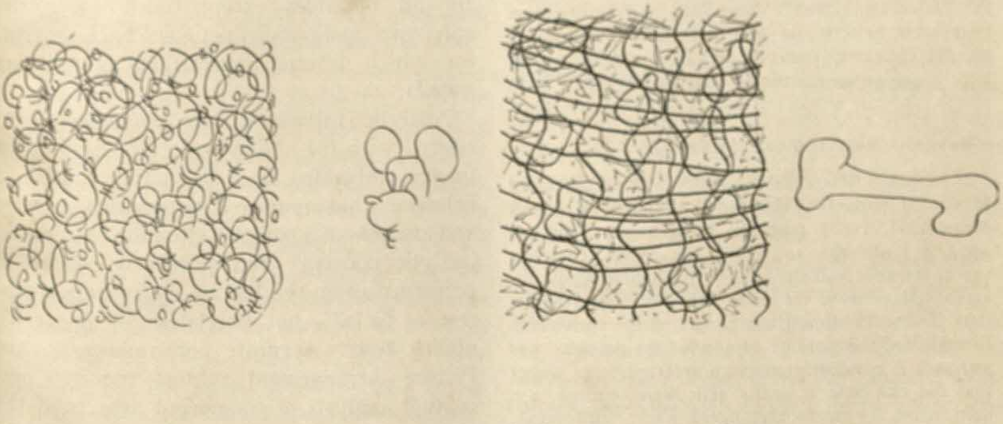


FIG. 3. Sample items from HFT.

Administration of Tests: Session II.

1. Paired-associate learning—The *S* learned four sets of four-word pairs. In each group of four-word pairs the stimulus words were nouns and the response word was either a verb, a coordinate noun, an adjective, or a noun that had a part-whole relationship to the stimulus noun. The four sets follow:

Set 1: Father-work, horse-chicken, ball-round, shirt-buttons

Set 2: Truck-wheels, knife-cuts, cup-glass, pear-yellow

Set 3: Bread-meat, pillow-soft, book-page, lion-roars

Set 4: Stick-long, bird-wing, rain-falls, pants-shoes

The *E* read each stimulus word and gave the child 10 seconds to answer. If no answer was forthcoming, *E* gave the child the correct answer. If *S* gave the wrong answer, *E* corrected him. The four words in a set were read to the child until he had mastered each set to a criterion of one correct trial (i.e., all four pairs in the set were correctly anticipated). This procedure was repeated until *S* had mastered all four sets. The variables scored were (*a*) the ease of learning each of the lists, and (*b*) differential difficulty in learning coordinate noun pairs when general learning ability was controlled statistically.

2. WISC—Four subtests of WISC were administered to *Ss*, two from the Verbal scale and two from the Performance scale. Vocabulary and Information tests were selected as the best indexes of verbal ability; Picture Completion and Picture Arrangement were regarded as related to analytic attitude. Each of the four raw scores was converted to a scale score.

Administration of Tests: Session III.

1. Modified Rorschach—Each *S* was administered 16 stimuli from the Modified Rorschach (Kagan & Moss, 1961, 1962) with the instructions, "Tell me what it looks like to you." The child's complete verbal response and the latency to his first significant verbalization were recorded. The average reaction time for the 16 stimuli was computed for each child. In addition, the number of nondistinct percepts was assessed. A nondistinct response referred to an object that did not have a clearly articulated outline (e.g. mud, paint, ice cream, dirt, cloud, water, stuff, lava). Interrater reliabilities (percentage agreement) for scoring this variable from protocols of children age 6–16 averaged 90%.[5]

Results

In the first analysis the number of analytic responses on CST was correlated with 13 variables derived from the other tests. Table 1 contains these correlations (product moment for all variables except coordinate verbs, antonyms, and nondistinct Rorschach percepts for which biserial correlations were computed).

TABLE 1

RELATION BETWEEN ANALYTIC CONCEPTS AND OTHER VARIABLES

Variable	Grade 1	Grade 2
Word association		
Noun—coordinate noun	.31*	.61***
Verb—coordinate verb ($r_{\text{bis.}}$)	−.01	.38*
Adjective—antonym ($r_{\text{bis.}}$)	.47**	.42**
Mean response time	−.09	.01
Hidden figures		
Weighted speed	.34*	.58***
Paired-associate learning		
Total errors	−.12	−.29
Errors in learning coordinate nouns	.13	−.20
WISC		
Information	.29	.19
Vocabulary	.06	−.08
Picture Completion	.22	.37**
Picture Arrangement	.58***	.55***
Ink blots		
Mean response time	−.09	.01
Indistinct response ($r_{\text{bis.}}$)	.01	−.05

* $p < .10$, two tailed
** $p < .05$, two tailed.
*** $p < .01$, two tailed.

Analytic responses were moderately correlated with the ability to locate the familiar figures embedded in a distracting context, a behavior that required systematic scanning and analysis of a complex stimulus. Moreover, analytic concepts were positively related to performance on the Picture Arrangement subtest of WISC, but unrelated to the verbal ability tests. Accurate performance on the Picture Arrangement subtest requires perceptual analysis of component aspects of the picture. For correct arrangement of the scenes into a coherent theme can only occur if the child attends to subtle cues such as posture of people and direction of travel of objects.

The expectation that syntactically congruent word associations to nouns, verbs, and adjectives would be related to analytic attitude was moderately supported, especially among second grade children. Analytic boys had a strong tendency to respond to nouns with coordinate nouns, to adjectives with antonyms, and to verbs with coordinate verbs. Coordinate nouns, verbs, and antonyms were positively correlated with each other. However, an analytic attitude was unrelated to ease of learning the noun-coordinate noun pairs in the paired-associate learning task.

---

[5] For a complete description of other tests given to these subjects see Rosman (1962).

The children who did not give coordinate noun responses produced either verb phrase completions to nouns (e.g., dog-bark), or adjectives describing nouns (e.g., dog-brown). In rare circumstances they produced words unrelated to the stimulus (e.g., dog-pipe, chair-bobby) or rhymes. Most *S*s produced an association that either matched temporal contiguity in language or an association that contained syntactic and semantic communalities with the stimulus. It is not surprising that the latter tendency increased with age, but it is surprising that this tendency was related to an analytic attitude, especially when analytic scores were independent of vocabulary size. The possibility that coordinate nouns or verbs were produced following reflection was not supported by study of the reaction time data. For coordinate responses or antonyms were not associated with longer response latencies on the word-association test.

Analytic Concepts and Impulsive Solution Attempts. It was suggested that a necessary antecedent of analytic concepts was a tendency to delay the reporting of solution hypotheses. One indirect measure of this tendency for delay was the occurrence of initially incorrect answers on HFT. As a first test of this hypothesis, children were classified as above or below the median on number of analytic responses on CST, and this dichotomized variable was related to the occurrence of one or more incorrect answers on HFT (for each grade separately). The resulting chi square was highly significant ($p < .01$), with 25 of the 31 analytic *S*s falling below the median on incorrect hypotheses, and 22 of the 31 low analytic *S*s falling above the median. It might be argued, however, that since analytic responses were positively correlated with HFT performance, low analytic subjects should have higher incorrect answer scores. In order to control for this relation, *S*s high or low on analytic concepts (based on median split) were matched on performance on HFT (i.e., in terms of solution times). Seven matched pairs of subjects were obtained among first grade boys, and six pairs among the second grade boys (total N of 26). Eleven of the 13 analytic *S*s were below the median on production of incorrect hypotheses, whereas 11 of the 13 less analytic *S*s were above the median ($p < .05$). Thus, even when ability to solve HFT was controlled, there was still a strong relation between analytic responses and inhibition of incorrect solution hypotheses.

Analytic Concepts and Indistinct Ink Blot Interpretations. There was no clear relation between analytic responses and indistinct ink blot interpretations when each grade was considered as a whole. However, when verbal facility was controlled by doing separate analyses for boys above or below the median on combined vocabulary and information scores, the expected negative relation between indistinct ink blot interpretations and analytic concepts was significant for boys with average or above average verbal facility (contingency coefficient = .46; $p < .05$), but not significant for boys with below average verbal ability. There was no relation between verbal facility and number of indistinct ink blot interpretations. The absence of a negative relation between analytic concepts and indistinct percepts for boys with low ability is attributable to the fact that many boys with below average verbal skills perseverated with simple form responses (e.g., bat, bird, butterfly) which were not scored as indistinct.

Response Time to Ink Blots. It will be recalled from Table 1 that there was no relation between speed of reaction to the ink blots and analytic concepts. When the relation between these dimensions was assessed separately for boys high and low on verbal facility the expected association between analytic concepts and long response latencies emerged only for boys with above average verbal ability (mean latencies of 7.6 seconds for analytic boys and 4.2 seconds for nonanalytic boys, $p < .05$). Low verbal facility impaired rapid conceptualization of the ambiguous ink blots, and there was a slight negative relationship between verbal facility and response latency.

In sum, the distinctness of the child's ink blot interpretations and the speed of his conceptualizations were influenced by both an analytic attitude and verbal facility. These two dimensions are orthogonal and, as will be seen, are critical determinants of the classificatory behavior of the child.

Summary

The results of this study suggest that an analytic attitude (*a*) appears in a variety of tasks with widely different intellectual requirements (i.e., articulation of ink blots; Picture Arrangement subtest of WISC; HFT performance); (*b*) is independent of vocabulary level; (*c*) is associated with the tendency to inhibit the impulsive reporting of inaccurate responses; and (*d*) is related to the production of word associations that, from a syntactic point of view, are developmentally more mature.

These findings extend and support the results of previous work. The association between an analytic attitude and inhibition of impulsive inaccurate responses is an important clue to understanding the significance of this response, and most of this monograph is devoted to study of the dimension of reflection impulsivity in situations in which *S* must choose among alternative solution hypotheses.

STUDY 2

It appears that analytic responses are produced following reflection over alternative responses. There are several indirect sources of evidence for this statement. First, reaction times to analytic concepts were noticeably longer than response times to other classes of concepts. In one group of second and third grade children (70 boys and 70 girls), the mean reaction time for analytic concepts was 5.8 seconds for boys and 6.9 seconds for girls. The reaction times for relational concepts (the most frequent response category) were 3.9 seconds for boys and 4.6 seconds for girls. This difference in reaction times to analytic versus relational responses was highly significant ($p < .01$). An earlier study (Kagan et al., 1963) reported a mean reaction time of 5.4 seconds for analytic and 4.0 seconds for relational concepts for sixth grade children ($p < .05$). It is interesting to note that the magnitude of the difference between response times for analytic and relational concepts was the same for the older and younger children. Analytic responses are preceded by a delay and it is suggested that during this delay the child is reflecting over alternative responses.

These results, together with those from the previous study (Study 1), suggested the wisdom of initiating an experimental study to test the hypothesis that a set to reflect would increase analytic concepts, a set to respond quickly would interfere with analytic responses. The experiment manipulated the occurrence of analytic responses through the use of instructions requesting the child either to delay before answering or to respond quickly. The three conceptual tests administered were:

1. CST
2. Design Recall Test (DRT). In this task a geometric design was presented for study, removed for 15 seconds, and then presented along with variants on the standard. The *S* had to select the correct design. Analytic performance was characterized by absence of incorrect guesses (i.e., the ability to select the correct answer on the first attempt). Figure 4 contains some sample items.[6]
3. HFT

It was expected that the instructional set to respond quickly would produce fewer analytic concepts and more errors on DRT and HFT than the instruction to reflect. As a control task, the Vocabulary scale from WISC was administered to all children under slow or fast instructions. It was assumed that correct reporting of the definition of a word would not be facilitated by reflection, and total vocabulary scores should be the same for all children.

Procedure

Subjects. The *S*s were 180 second grade children (90 boys and 90 girls). The school had divided the children into three levels of reading ability (based on group test data), and within each of these three groups children were paired on sex, intelligence (California Test of Mental Maturity or CTMM), and education of parents. There was a high relation between reading ability, on the one hand, and both intelligence and parental education level on the other. One member of each pair was placed in a fast instruction group; the second was placed in a slow instruc-

---

[6] These stimuli were designed originally by Eleanor Gibson and her colleagues at Cornell University.

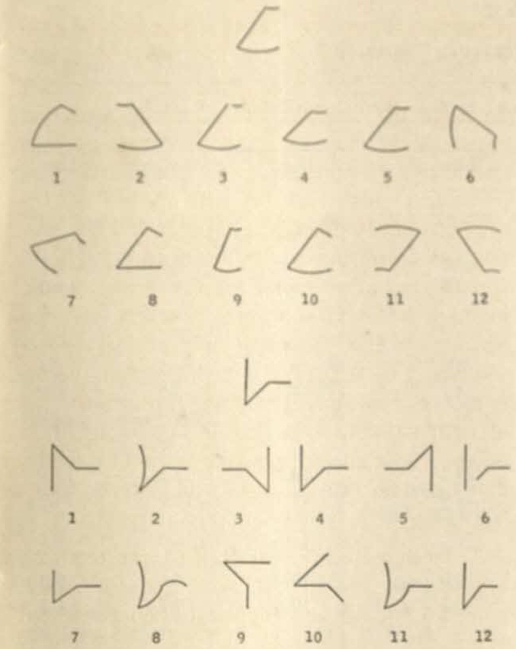


FIG. 4. Sample items from DRT.

tion group. The study involved a 2×2×3 design with sex, instructional set, and general intelligence (three levels) as discrete variables with 15 subjects in each of 12 cells.

Each child was administered four tests in one 60-minute session. Initially, 120 subjects were administered CST, DRT, HFT, and WISC Vocabulary in one session. This session was too long, and the last 60 subjects were not given HFT. Thus, for the analyses that involve HFT there were only 10 subjects in each of 12 cells.

Administration of Tests.

1. CST—The 30-item CST was administered with two different instructions. The instructions for the *fast* group were:

> I have some cards here with three pictures on them like this one. I want you to pick out two pictures that are alike or go together in some way. Try to get the answer to each card as fast as you can. What two pictures are alike or go together here? Answer as fast as you can.

The instructions to the *slow* group were similar, but *E* added,

> Take your time and think about your answer before telling it to me. What are the two pictures that are alike or go together here? Take your time and think about your answer.

2. DRT—The *S* was first shown a simple geometric design for 5 seconds and told to remember it. The standard was then removed from *S*'s view for 15 seconds after which an array of 8, 9, or 10 similar stimuli was presented. The *S* was instructed to select the one design that was identical to the standard. Each *S* was first given two practice items followed by presentation of 12 test items. Instructions for the fast group included the exhortation,

> Try to pick out the right one as fast as you can.

The instructions for the slow group emphasized reflection,

> When I show you the group of designs take your time before answering so that you are sure that you have picked out the right design.

3. HFT—This was the same test that was administered in Study 1. The instructions for the fast group emphasized speed of response (i.e., "Can you find the fish? Point to the fish as fast as you can."). The instructions for the slow group emphasized reflection (i.e., "Take your time in finding the fish. Make sure you are right before pointing to it.").

4. *Vocabulary*—Each S was administered the Vocabulary scale of WISC. The *S*s in the slow instruction group were told to "think about the answer" before responding. The *S*s in the fast instruction group were told to respond "as quickly as possible."

Results

Table 2 presents the means for the major variables for each of the 12 experimental cells.

The major hypotheses of the investigation were verified, for with slow instructions the children were more likely to (*a*) produce more analytic responses on CST, (*b*) select initially the correct figure on DRT, (*c*) select initially the correct hidden figure (p's $< .001$, $< .01$, $< .001$ by t test, respectively). Analyses of variance for each of these three responses are presented in Table 3.

As predicted, vocabulary scores were not affected by the different instructional sets; the mean scores for the fast and slow groups were identical. The response time to each of the four tasks was clearly influenced by the instruction to delay or to respond quickly $p < .01$ for each task). Thus, the increase in analytic behavior under slow instructions was accompanied by increased delays before responding. Moreover, even within a group given the same instructional set, there was a positive relation between response time and analytic performance. Average correlations for each of the three verbal ability groups (based on z' transformations) were computed among the following five variables: (*a*) response time to CST, (*b*) analytic responses on

TABLE 2

MEAN SCORES FOR CONCEPTUAL VARIABLES

Variable	Instructions (2 levels)—verbal intelligence (3 levels)					
	Fast			Slow		
	High	Medium	Low	High	Medium	Low
CST						
Analytic responses						
Boys	2.3	3.9	1.5	4.6	6.6	3.7
Girls	2.8	3.3	3.5	6.4	4.1	4.0
DRT						
Correct first attempt						
Boys	6.7	5.1	4.4	6.9	6.7	5.4
Girls	6.1	4.5	4.2	7.4	5.7	4.5
DRT						
Errors						
Boys	12.6	18.7	21.1	10.2	12.1	19.5
Girls	13.3	22.9	22.4	11.9	16.9	24.9
Hidden figures						
Correct first attempt						
Boys	9.5	7.6	8.0	10.0	9.6	9.3
Girls	8.8	9.1	8.2	10.6	9.5	8.4
WISC Vocabulary						
Boys	25.9	21.3	21.5	27.5	23.0	15.2
Girls	23.8	17.3	18.3	22.7	21.4	18.1
CST						
Response time						
Boys	3.5	3.0	3.8	6.7	8.5	7.3
Girls	3.0	3.9	3.6	7.5	6.0	6.7
DRT						
Response time						
Boys	4.6	4.3	4.5	8.0	9.0	7.0
Girls	4.5	4.8	4.4	9.6	7.5	7.0
Hidden figures						
Response time						
Boys	10.3	8.6	6.4	13.2	11.3	10.1
Girls	9.7	7.4	6.0	12.2	10.3	10.0
Vocabulary						
Response time						
Boys	3.2	2.4	2.4	4.5	5.0	4.9
Girls	3.1	5.0	2.6	5.1	4.8	5.9

CST, (*c*) response time to first response on DRT, (*d*) number correct on first attempt on DRT, and (*e*) intelligence. The data for HFT were omitted because the sample size was only 10 for each cell.[7] There was a positive relation between number of analytic responses on CST and reaction time for the 30-item test (r's = .40 and .42 for boys and girls in the fast instruction groups, $p < .01$; r's = .25 and .44 for boys and girls in the slow instruction groups, $p < .01$). Similarly, there was a positive relation between number of designs correct on first attempt on DRT and reaction time to first selection ($r = .43$, $p < .01$; and $r = .32$, $p < .05$ for boys and girls in the fast group; r's = .27 and .26 for boys and girls in the slow group). The relevance of delay for an analytic attitude is supported by these positive relations between length of delay be-

---

[7] Complete tables of means, standard deviations, or correlation matrices can be obtained by writing to the author.

TABLE 3

SUMMARY TABLES FOR ANALYSES OF VARIANCE

Source	*df*	*SS*	*MS*	*F*
Analytic responses on CST				
Instructions (I)	1	180	180.0	16.4***
Sex (S)	1	2	2.0	
Verbal ability (V)	2	73	36.5	3.3*
I×S	1	6	6.0	
I×V	2	22	11.0	
V×S	2	36	18.0	
I×S×V	2	80	40.0	3.4*
Within (error)	168	1846	11.0	
Number correct DRT first attempt				
I	1	39	39.0	9.30**
S	1	10	10.0	
V	2	137	68.5	16.10***
I×S	1	0	0	
I×V	2	3	1.5	
V×S	2	4	2.0	
I×S×V	2	7	3.5	
Within (error)	168	713	4.2	
Number correct HFT first attempt				
I	1	32	32.0	13.9***
S	1	1	1.0	
V	2	32	16.0	7.0**
I×S	1	1	1.0	
I×V	2	1	1.0	
V×S	2	5	2.5	
I×S×V	2	13	6.5	
Within (error)	108	243	2.3	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

fore responding and either number of analytic responses or number of designs recognized on first attempt.

In a final analysis the matrix of intercorrelations for the 12 major variables was computed. The three intelligence groups were collapsed, and correlation coefficients were based on groups of 45 boys and 45 girls in either the slow or fast instruction group. Coefficients that involved HFT were based on 30 subjects. In this analysis a differentiation was made between major and minor errors on DRT. The most frequent error was the selection of a variant that only differed in size or proportion from the standard (Variants 4 and 9 in top design in Figure 4). The other variants all involved more obvious distortions of the standard, e.g., changes in the curvature of designs, changes in the orientation of the design, breaks in the continuity of a line (Variants 1, 2, 3, 6, 7, 8, 10, 11, and 12 in Figure 4). In Tables 4 and 5 major errors refer to the more obvious alterations which occurred less frequently; minor errors refer to the more common error in which the child selected a stimulus differing only in size or proportion from the original. The coefficients appear in Tables 4 and 5.

Analytic concepts on CST showed a negative relation to major errors on DRT (r's = $-.16$, $-.25$ for *S*s in fast group; r's = $-.11$, $-.41$ for *S*s in slow group; average $r = -.24$, $p < .01$). Analytic responses were, as expected, not highly related to standard IQ score (r ranged from $-.16$ to $.16$). When the analysis was restricted to children given either fast or slow instructions, there was still a strong, positive relation between number of analytic concepts on CST and response time across all 30 items (r's = .26, .46, .32, .39; p's < .10, < .01, < .05, < .01). Further, there was intraindividual consistency in the tendency to have long response times on both CST and DRT (r's = .43, .53, .49, .36; $p < .05$ for all coefficients). For children under slow instructions there was even a positive correlation between response times on HFT and CST ($r = .33$, $p < .10$; $r = .59$, $p < .001$). Finally, response times to HFT and DRT were positively correlated (r's = .31, .10, .39, and .26; the boys' coefficients of .31 and .39 were significant at $p < .10$ and < .05, respectively). It is to be noted that response times to CST, DRT, and HFT were independent of vocabulary scores (11 or 12 coefficients fell between $-.16$ and .16). This finding is replicated in other studies to be summarized.

Unlike analytic concepts on CST, accurate performance on DRT had a moderate relation to IQ score on CTMM. This could be due to several factors. First, DRT involved memory. The ability of a second grade child to retain a clear image of a meaningless geometric design for 15 seconds may be facilitated by the availability of language labels that enable the child to classify the design. The stimuli used in the test were relatively meaningless forms and the child with a rich vocabulary may have been more likely to

TABLE 4

INTERCORRELATIONS AMONG COGNITIVE VARIABLES FOR CHILDREN UNDER FAST INSTRUCTIONS

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. CST, Analytic responses	—	.26*	—.16	—.21	.25*	.10	.04	—.13	.14	—.10	.22	—.16
2. CST, Response time	.46***	—	—.08	—.21	.13	.43***	.18	—.08	.15	—.07	—.15	—.22
3. DRT, Major errors	—.25*	—.16	—	.45***	—.72****	—.27*	—.48***	.46***	—.23	—.13	—.46***	—.54****
4. DRT, Minor errors	—.18	.03	.46***	—	—.70****	—.43***	—.34*	.38**	—.07	—.08	—.17	—.21
5. DRT, Correct first attempt	.22	.04	—.57****	—.62****	—	.38***	.55***	—.52***	.30*	.02	.40***	.52****
6. DRT, Response time	.36**	.53****	—.36**	—.22	.33**	—	.26	—.22	.31*	—.10	—.11	.13
7. HFT, Correct first attempt	—.23	—.01	—.05	—.06	.35*	.04	—	—.83****	.26	.28	.26	.31*
8. HFT, Total errors	.11	—.17	.27	.24	—.39**	—.20	—.87****	—	—.44**	—.08	—.31*	—.29
9. HFT, Response time	—.01	.03	.00	—.12	.26	.10	.18	—.17	—	—.62****	.13	.11
10. Speed of solution	—.41**	—.38**	—.09	—.06	.22	—.39**	.32*	—.23	—.36**	—	—.04	.01
11. WISC, Vocabulary	.10	—.09	—.37**	—.15	.18	—.13	—.09	—.05	.16	.24	—	.31**
12. IQ (CTMM)	.05	—.25	—.38***	—.38***	.32**	—.15	—.14	.08	.37**	—.05	.54****	—

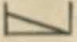
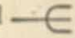
Note.—Males to the right and above the diagonal; females to the left and below.
* $p < .10$, two tailed.
** $p < .05$, two tailed.
*** $p < .01$, two tailed.
**** $p < .001$, two tailed.

TABLE 5

INTERCORRELATIONS AMONG COGNITIVE VARIABLES FOR CHILDREN UNDER SLOW INSTRUCTIONS

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. CST, Analytic responses	—	.32**	—.11	—.07	—.10	.16	—.08	—.12	.25	—.32*	.15	.16
2. CST, Response time	.39***	—	—.09	—.03	—.02	.49****	—.13	.21	.33*	—.46***	—.16	.05
3. DRT, Major errors	—.41***	—.16	—	.59****	—.76****	—.22	—.50***	.52***	—.13	—.11	—.33**	—.45***
4. DRT, Minor errors	—.37**	—.19	.72****	—	—.73****	—.38***	—.29	.58****	—.40**	.06	—.22	—.31**
5. DRT, Correct first attempt	.32**	.20	—.79****	—.84****	—	.33**	.42**	—.55***	.28	.00	.15	.35**
6. DRT, Response time	.28	.36**	—.27*	—.36**	.37**	—	.01	—.26	.39**	—.50***	.13	.24
7. HFT, Correct first attempt	.37**	.48***	—.46***	—.43**	.39**	.40**	—	—.67****	—.11	.40**	.37**	.04
8. HFT, Total errors	—.35*	—.54***	.37**	.36**	—.39**	—.33*	—.89****	—	.00	—.19	—.30*	—.10
9. HFT, Response time	.41**	.59****	—.16	—.02	.11	.26	.05	—.13	—	—.76****	—.11	.07
10. Speed of solution	—.14	—.47***	—.07	—.12	.06	—.29	.12	.00	—.75****	—	.25	—.02
11. WISC, Vocabulary	.34**	.12	—.37**	—.50****	.43***	.36**	.52***	—.43**	.25	.01	—	.50****
12. IQ (CTMM)	.07	.13	—.32**	—.37**	.45***	.35**	.40**	—.35*	.16	—.06	.44***	—

Note.—Males to the right and above the diagonal; females to the left and below.
* $p < .10$, two tailed.
** $p < .05$, two tailed.
*** $p < .01$, two tailed.
**** $p < .001$, two tailed.

apply a label to the design that facilitated its subsequent retention. For example, the design  was called by some children a "tilted glass of milk"; and  was sometimes called a "pitchfork." The child who was quick to apply such labels might be expected to perform better on this test than one who did not use this verbal device. Finally, DRT resembled a school task more than CST. It was relatively difficult, and there was clearly a right answer. The child of average intelligence may not have invested intense interest in the task because of minimal desire to perform competently or minimal confidence in his ability to perform adequately.

The HFT was too easy for this group, for over 90% of the children solved the designs within 20 seconds. Thus, the critical measure on this task was not solution time, but the *S*'s approach to the task (i.e., reaction time to first guess and number of errors before achieving solution). It will be recalled that children who achieved many initially correct solutions on HFT also achieved many initially correct solutions on DRT, indicating that reflection over alternative responses was consistent across the two tests.

Discussion

These data argue for the significance of a dimension of reflection versus impulsivity in problem situations where many solution hypotheses are available simultaneously, and the child has to evaluate the differential adequacy of each possibility. Some children act first and discover later if they were correct; others reflect before responding, eliminating potentially incorrect answers mentally.

The importance of this subtle cognitive process may not be fully appreciated. An impulsive approach to problem solving is likely to be associated with the execution of solution hypotheses that are incorrect. The impulsive child is apt to act upon his initial hunch with minimal reflection, carry through a false hypothesis mentally, and offer an answer without critical examination of its potential accuracy. The anxiety resulting from repeated failure, as a consequent of impulsive reporting, could lead to generalized expectations of failure and withdrawal of involvement from intellectual tasks.

STUDY 3

The previous study suggested that the tendency to inhibit impulsive, inaccurate answers to problems with alternative solution possibilities was consistent across tasks and amenable to experimental manipulation. This investigation is a partial replication and extension of the previous one. The primary purpose was to assess the relation between analytic concepts on CST and impulsive errors on DRT under a "normal" testing situation (i.e., where the child was not urged to respond either quickly or slowly). A second purpose was to test the hypothesis that children with high analytic scores would reflect over their cognitive products and examine them for completeness and accuracy. Children were asked to draw an object with which they were all equally familiar—a human face—and their products were examined for differences in completeness.[8] The analytic *S*'s product was expected to contain more relevant elements than that of the less analytic child. A final purpose was to replicate an earlier unpublished finding of a low positive relation between the tendency to produce analytic concepts and the ability to prolong the simple motor task of drawing a straight line.

Procedure

Subjects. The *S*s were 72 boys (35 in Grade 2 and 37 in Grade 3) and 63 girls (34 in Grade 2 and 29 in Grade 3) whose ages ranged from 6-10 to 9-6 with a median age of 7-10. Of the total group of 135, 112 were Caucasian, 20 were Negro, and 4 were of Oriental background. Most of the boys (77%) were seen in Study 1.

The educational level of the parents ranged from part high school to professional training; the modal educational level was high school graduate. All *S*s were seen individually for two 60-minute sessions by a young adult of opposite sex to the child. The two sessions were separated by a period of 9 weeks.

Administration of Tests: Session I.

1. CST—The 30-item CST was administered to each child with standard instructions. The two vari-

---

[8] This idea was suggested by G. A. Miller.

ables coded were (*a*) analytic responses and (*b*) mean response time to all 30 items.[9]

2. Delayed Recall Designs Test (DRT I)—This task was administered as in Study 2, but *S* was told neither to respond quickly nor slowly. He was left to respond according to his own natural tempo. As in Study 2 the most frequent error was to the variant in which the stimulus was identical in shape and orientation to the original, but the proportion was altered in a minor way (called a minor error). All of the other variations on the standard included changes either in a part of the stimulus or in its orientation (major errors). Two different error scores were computed as well as mean response time across all 12 items, and number of items for which *S*'s first selection was correct.

3. Motor inhibition task: Draw-a-line (DAL)—The *S* was asked to draw a straight line as slowly as possible. The *E* first drew part of a line very slowly and then asked *S* to

> Draw a line like this but go as slowly as you can. You can't stop and you can't take your pencil off the paper. I'll tell you when to start. O.K. Go.

After *S* had drawn the line *E* said,

> Now I want you to do this once more. But go even slower. Do the same thing but try to make the line even slower.

The variables analyzed were the time taken to complete the line for each trial (DAL I, DAL II).

4. Verbal fluency (VF I)—The child was asked to think of as many things as he could that were round, and each *S* was given 2 minutes to respond. The only variable of interest was the number of appropriate items named. This task was included as a control task; it was expected that the fluency score would be unrelated to measures of analytic attitude.

Administration of Tests: Session II (9 weeks later).

1. Picture Discrimination Test (PDT)—The *S* was shown two similar pictures and asked to detect the difference between them. The items were taken from Meier's Art Preference Test. After four practice items a set of 15 test items was administered, with a time limit of 2 minutes allowed for each item If *S*'s initial answer was incorrect *E* said,

> That is not the most important difference. Can you find another way in which the pictures differ?

The *E* allowed the child to guess as many times as he wished within the 2-minute period. Figure 5 presents two sample items from the test.

The major variables coded were (*a*) number of incorrect solution hypotheses, (*b*) number of items for which *S*'s first solution hypothesis was correct, and (*c*) average time to solution across all 15 items.

After PDT, *S* was administered a special task designed to assess his perception of school. These data will not be reported here.

2. Design Recall Test (DRT II)—The DRT was administered again with the same instructions used in the first session, but the items were presented in a different order.

3. Verbal fluency (VF II)—Each *S* was given a second verbal fluency task in which he was asked to report all things that he could think of that were square in a 2-minute period.

4. Draw-a-face (DAF)—The *S* was requested to draw a human face as if he were looking at it. The order of appearance of each of the face parts was recorded by *E*. The major variables scored were the number of parts included (i.e., eyes, nose, mouth, ears, hair, wrinkles, etc.) and the order of appearance of each part.

In addition, language and nonlanguage IQ scores based on CTMM were available for 92% of the children. This test had been group administered to *S*'s prior to the two test sessions.

Results

Sex Differences. There were significant sex differences in mean scores for some of the variables (see Footnote 7). Boys, in comparison to girls, produced more analytic responses ($p < .001$), higher fluency scores ($p < .01$), and more errors on PDT ($p < .05$). Grade 3 boys produced less major errors on both DRT I and II than Grade 3 girls ($p < .05$) and took a significantly longer time to draw the straight line than third grade girls ($p < .01$ for Trial 1 and $p < .05$ for Trial 2). Girls, on the other hand, had higher scores than boys on average response latency to CST ($p < .01$), average latency to first response on DRT I and II ($p < .01$ and $p < .001$, respectively), and number of face parts ($p < .05$). In sum, boys produced more analytic concepts and fewer recognition errors on DRT, despite the fact that girls had longer response times to CST and DRT. The longer response times among girls were no guarantee that they would make fewer errors on these visual analysis tasks.

Age Differences. Third grade boys made fewer major errors on DRT than second grade boys for both sessions ($p < .05$); but third grade girls made slightly more errors than second grade girls. This latter difference just missed statistical significance. Fluency scores and number of face parts drawn increased with age for both sexes ($p < .01$ and $p < .05$).

Intercorrelations among the Variables. In-

---

[9] The scores for the 55 boys for whom this was the second administration of CST were based on their performance on this second test.

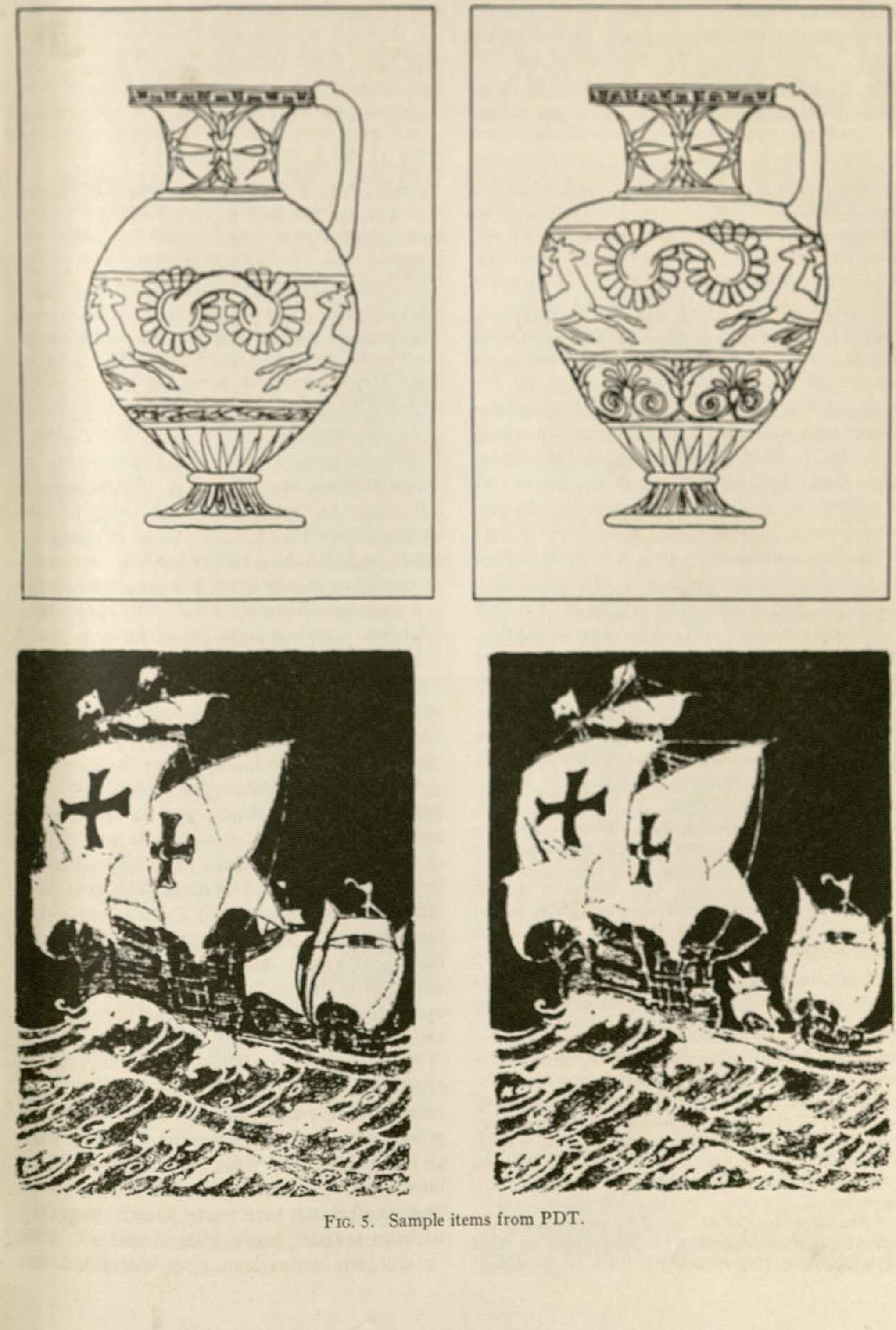


FIG. 5. Sample items from PDT.

tercorrelations among these variables (scores were transformed to McCall *T* scores) were computed for the four sex-grade groups separately (see Footnote 7). There are several sets of provocative relations contained in these coefficients.

Consistency of Response Time Scores. The child's tendency to display long or short response times on CST, DRT I, and DRT II was highly consistent. The intercorrelations among these three response times were all positive (*r* ranged from .21 to .77 with an average *r* of .49), and 10 of the 12 coefficients were statistically significant at .05 or better. The coefficients relating response times on DRT I and II were .77 and .29 for boys ($p < .001$ and $p < .10$), and .50 and .66 for girls ($p < .01$ and $p < .001$). These correlations were similar in magnitude to those reported in Study 2, and support the conclusion that the tendency to delay before responding in a situation with many alternatives shows intraindividual consistency across situations and stability over a 9-week period.

Consistency of Error Scores. The reliability of major and minor error scores on the two administrations of DRT was unusually high. All eight coefficients were significant and ranged from .29 to .84 with an average *r* of .58. The reliability of the major error score was slightly greater than the minor error score. The unusually high coefficient of .84 for major errors occurred among Grade 3 girls. Subsequent investigations on an independent sample of 70 second grade children yielded a reliability coefficient of .55 for DRT errors with a test-retest interval of 7 months. The intratest reliability (odd-even) for total errors was .77 for the present group of second and third graders.

A tendency to offer incorrect solution hypotheses on PDT was positively related to major errors on both administrations of DRT for second grade *S*s, but not for third graders. For the Grade 2 *S*s the four correlations ranged from .52 to .73 (all significant at $p <$ .01). It will be recalled from Study 2 that errors on DRT were positively related to errors on HFT (*r* ranged from .27 to .58).

Relation between Response Time and Errors. There was a consistent negative relation between response times on CST or DRT and error scores on DRT I, DRT II, and PDT. As might be expected the relationship was highest when response times and error data came from the same test. For example, the relationship between major errors and response time on DRT I was −.45, −.29, −.42, and −.55 for the four groups. For DRT II the coefficients ranged from −.09 to −.48 with an average *r* of −.28. There was typically a strong negative relationship between response time on one task and errors on a second task. To illustrate, response time to CST was negatively associated with major errors on DRT I for three of the four groups (*r*'s = −.42, −.26 for boys; *r*'s = .11, −.28 for girls). Similarly, long response times to DRT I were negatively related to major errors on DRT II for three of the four groups (*r*'s = −.44, .00 for boys; *r*'s = −.32, −.51 for girls). Indeed, for second grade boys the relation between response time and major errors on DRT I ($r = -.45$) was no better than the relation between response time on DRT I and major errors on DRT II 9 weeks later ($r = -.44$). These data are consonant with those of Study 2, for response times to DRT were negatively related to errors on both DRT and HFT.

In sum, these children showed a *stable tendency toward delay versus impulsivity in reporting solution hypotheses,* and this tendency was *consistently related to low versus high error scores on DRT.*

Relationships with Analytic Responses. The index of analytic attitude was the number of analytic responses on CST. The pattern of correlations between analytic concepts and the other variables was in a direction that might be expected. Analytic responses were associated with slightly longer response times to CST (*r*'s = .21 and .27 for boys; *r*'s = .07 and .32 for girls), and with longer response times to the DRT I for second grade girls ($r = .46$; $p < .01$), and for third grade boys ($r = .32$; $p < .05$). The correlations for the other two groups were .07 and .10. Analytic responses were also associated with low major error scores on DRT I and II for all groups ($r = -.36, p < .05$ and $r = -.48, p < .01$ for boys; *r*'s= −.27 and −.50; $p < .01$ for girls). These correlations were somewhat higher than the corresponding coefficients in Study 2 where the subjects were given a set to delay or respond quickly. Finally, analytic responses were associated with low error scores on PDT for three of the four groups ($r = -.33, p < .05$ and $r = .14$ for boys; $r = -.42$, $p < .05$ and $r = -.48$, $p < .01$ for girls). There was a suggestive positive relation between analytic concepts and number of face parts drawn (average $r = .17$; $p < .05$). Analytic concepts were, as predicted, independent of verbal fluency scores, but moderately related to language IQ (r's $=.23$ and $.35$ for boys; r's $=.46$ and $.21$ for girls). The magnitudes of these relationships between group administered test scores and analytic concepts were the highest obtained in any of the studies during the last 4 years. Since group administered IQ tests have a multiple-choice format it is not unlikely that an analytic attitude could facilitate performance on this kind of instrument. However, when language IQ was partialed out through a regression analysis, the magnitude of the relation between analytic concepts and the other variables did not change to an appreciable degree. For example, analytic concepts remained negatively related to major errors on DRT I and II ($r = -.30$, $p < .10$ and $r = -.54$, $p < .01$ for boys; $r = -.53$, $p < .001$ and $r = -.23$ for girls).

Significance of Face Parts. All children drew the eyes, nose, and mouth, and differences in scores reflect the addition of attributes such as ears, hair, eyebrows, eyelashes, wrinkles, and teeth. It is suggested that one of the primary factors responsible for high scores on this task is the tendency to reflect on what a face is like while executing the drawing or after it is completed. The child who reflects on the adequacy of his production is apt to include more attributes than one who impulsively draws what first occurs to him and fails to reflect on the completeness of his product. These ideas are supported, to some degree, by the correlation between the number of face parts drawn and other variables. High scoring boys had longer reaction times to CST (r's $= .32$ and $.35$) and DRT (r's $= .29$ and $.29$) and made fewer minor errors on the DRT (r's $= -.47$ and $-.42$). There was no relationship between number of face parts drawn and IQ. For third grade girls, number of face parts was related to long response times on CST ($r = .49$) and, as with boys, was independent of IQ score.[10]

DAL. The ability to prolong the drawing of a straght line was stable across the two trials (r's $= .80, .87, .89, .81$), orthogonal to language IQ, and most clearly related to low error scores on DRT. For example, time to draw the line on Trial 1 was negatively related to minor errors on DRT I (r's $= -.44$ and $-.39$, for boys; r's $= .03$ and $-.20$ for girls) and to major errors on DRT I (r's $= -.54, .05, -.18, -.27$).

Fluency. The verbal fluency scores were also relatively stable across the two sessions (r's $= .63, .69, .29, .52$), but were generally unrelated to the other variables. The direction of the relationships favored a slim association between fluency and fast response times. To illustrate, the relationship between fluency for round objects and response time on DRT I was $-.11$ and $-.33$ for boys; $-.17$ and $.12$ for girls.

Discussion

These results, together with those from Study 2, provide persuasive evidence for the postulation of a construct of reflection versus impulsivity in offering solution hypotheses when several responses are available simultaneously. The purest measure of this construct is response time to first selection in tests like DRT. The relation between analytic concepts and response time or errors on DRT was only moderate. Thus, the determinants of an analytic concept must involve variables other than active reflection upon alternatives. One of these additional variables will be discussed in a later study.

---

[10] There was a significant age difference in the order in which the hair was drawn. Third grade children were not only more likely to include the hair than second graders ($p < .05$), but they were apt to draw it early (i.e., the first or second item produced). For the boys 33% of the third graders and 11% of the second graders drew the hair as the first or second component of the face. For girls 62% of the third graders and 29% of the second graders drew the hair early. The resultant chi square for the pooled sexes was 10.20 ($p < .01$). One interpretation of the increased salience of the hair with age assumes that type of haircut is a critical external criterion for sex-role differentiation. The age difference in the inclusion of the hair early in the drawing may reflect a heightened concern with sex-role differentiation in third grade children. It is possible that the order in which elements of a schema are expressed (in verbal or graphic form) might provide an indirect index of the individual's preoccupation with that element.

Although this line of inquiry began with CST, the results of several studies have led to a concern with a dimension (i.e., reflection-impulsivity) for which CST is not the best measuring device. An analytic response is complex and multiply determined, and it seems wise to concentrate on each of its more basic determinants. It is not uncommon that a new instrument acts as a wedge to uncover significant phenomena which in turn dictate the discarding of the original instrument. The phrase analytic attitude in this monograph shall refer specifically to analytic concepts on CST; whereas, the reflection-impulsivity variable refers to response time, when it is understood that long response times are associated with low error scores.

This study also yielded other ideas that warrant systematic study. The low, but suggestive, negative relation between fluency and reflection is intuitively reasonable. The child who pauses to evaluate the validity of each hypothesis should produce less volume in a brainstorming task like verbal fluency. The meaning of a fluency score, as suggested by the writing of Guilford (1959) and Getzels and Jackson (1962), is consonant with these data.

The fact that reflective *S*s drew more face parts has methodological consequences for investigators who use children's drawings as an index of the child's perception or mental constructions. Piaget and his colleagues (Inhelder and Piaget, 1958) occasionally use a child's drawings as a measure of his conception of an event. The Geneva investigators interpret developmental differences in the child's graphic productions as isomorphic with the child's understanding of an event. The present data indicate both age and individual differences in the accuracy and completeness of the reproduction of a face, an object for which the child has complete familiarity. When the child omits ears or hair in his drawing it is more reasonable to assume that he failed to reproduce these components as a result of inadequate reflection, rather than conclude that he does not know that people have hair or ears.

The *S*s in this study recently viewed six different line drawings, presented tachistoscopically, a minimum of 18 times at increasing exposures. The child reported what he saw after each exposure. After the last exposure *S* was told to look at the picture once more (at a 3-second exposure) and to draw it. It was not uncommon for a child to omit completely from his drawing an aspect of a scene he had described accurately a few seconds earlier. Moreover, there was a positive relation between completeness of the six drawings and both analytic concepts (average $r = .28$, $p < .01$) and number of face parts drawn (average $r = .29$, $p < .01$), and a negative relationship with errors on DRT II (average $r = -.30$, $p < .01$).

These additional data strengthen the earlier statement that differences in cognitive products should not always be attributed to differential knowledge, but may be a consequence of differences in the maturity of the response systems the child uses to communicate to *E*, be it a verbal explanation, a drawing, or a finger pointing to a picture. Psychologists should be more cautious about glibly attributing developmental differences in cognitive performance to differences in cognitive structures when it is likely that both classification and response processes are also changing systematically with age.

STUDY 4

The data summarized in the first three studies suggest the role of reflection impulsivity in facilitating analytic concepts on CST. A second dimension implied by the phrase "analytic attitude" is analysis (or fractionation) of visual stimuli. For an analytic concept on CST is a grouping based on similarity in *component* parts of the total stimulus. One way to test the notion that analytic concepts are produced by children who are predisposed to analyze stimuli is to require the child to learn a label to a new stimulus, and to assess, subsequently, the child's ability to apply this label correctly to component parts of the original stimulus.

Results from an early study (Kagan et al., 1963) revealed that fourth graders who produced many analytic concepts were more likely than less analytic children to attach a nonsense syllable to discrete parts of a geometric design. Briefly, 25 boys and 25 girls were administered a simple paired-associate learning task in which the child was required to learn four different nonsense syllables to each of four geometric designs. Figure 6 (bottom section) illustrates these designs.

After reaching a criterion of eight successive correct responses *S* was shown the figure and ground components separately (top half of Figure 6) and asked to apply the correct syllable. The girls made significantly more errors than boys when the figural component was presented alone (Ms = 3.0 versus 1.7, $p < .01$). Further, the boys and girls who produced many analytic concepts on CST were more likely to label the figure correctly when it was presented alone than less analytic children. The correlation between analytic concepts and frequency of correct labeling of the figure elements was .32 ($p < .05$).

Replication

In order to assess the generality of this result, a replication was initiated. A second purpose was to replicate the relation of analytic attitude to two variables already studied—the ability to detect differences between highly similar stimuli (using PDT) and the ability to prolong the drawing of a straight line.

Procedure

Subjects. The *S*s were 57 fourth grade boys; *E* was a female college student.

Administration of Tests.

1 CST—Each *S* was given CST with standard instructions.

2. Visual analysis—Each *S* was administered the four pairs of geometric designs illustrated in Figure 6, and asked to learn a nonsense syllable to each design until he reached a criterion of eight consecutive correct responses. After reaching criterion each *S* was shown the figure and ground components separately (eight stimuli in the transfer set) and applied the syllable he believed was correct.

3. PDT—This test contained 19 pairs of pictures which differed in a subtle aspect (see Figure 5 for illustrations). The instructions to the child were "to find the most important difference between them." The *S* was allowed 2 minutes to find the difference between them and was not told if his answer was correct or incorrect.

4. Second transfer trial: Visual analysis—Each *S* was again presented with the eight stimuli (in a different order) that contained the separate figure and ground elements.

5. DAL—Each *S* was asked to draw a straight line as slowly as possible. After the first trial, *E* told *S* to, "do it again, even slower, more slowly this time." The *S* then repeated the task.

6. Intelligence—Scores on the Pintner-Cunningham IQ test (group administered) were available for 43 of the 57 boys.

Results

Table 6 contains the correlations between analytic concepts on CST and each of the other major variables.

Analytic responses were associated with correct recognition of the figural element of the original design. There was no relationship between analytic concepts and either correct labeling of the background or ease of learning the association between nonsense syllable and the whole design.

Analytic boys were also better able to detect the differences between the two similar pictures on PDT. This study was completed before the results of Study 3 were analyzed, and at the time the importance of obtaining number of incorrect hypotheses on PDT was not recognized. Finally, there was only a slight relation, but in the predicted direction, between analytic concepts and the ability to prolong the drawing of the straight line. As

TABLE 6

RELATION BETWEEN ANALYTIC CONCEPTS AND OTHER VARIABLES

Variable	Correlation
Visual analysis,	
Trial of attainment	.14
Visual analysis,	
Figural component correct	.26**
Visual analysis,	
Background component correct	−.09
PDT, Failures	−.41***
DAL speed,	
Trial I	−.13
Trial II	−.09
Trials I and II (average)	−.25*
Intelligence ($N = 43$)	.24

Note.—$N = 57$ fourth grade boys.
* $p < .10$, two tailed.
** $p < .05$, two tailed.
*** $p < .01$, two tailed.

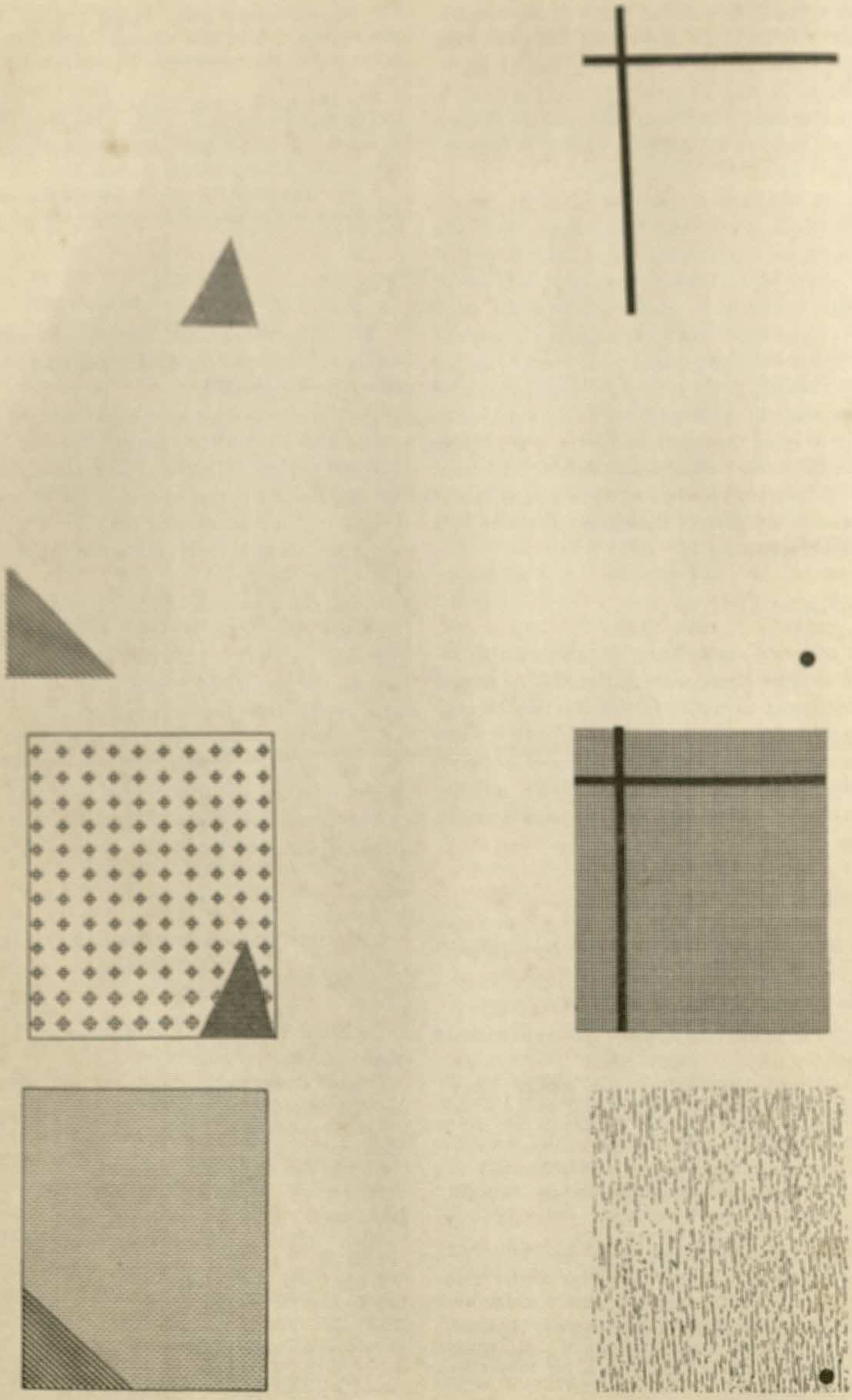


FIG. 6. Stimuli used in figure-ground learning task.

in most of the previous studies, there was no strong relation between analytic concepts and IQ score.

Discussion

These data, together with those from Studies 1, 2, and 3, suggest that analytic concepts are related to two distinct variables: the tendency to analyze a stimulus into its elemental components and the tendency to inhibit impulsive answers. The joint influence of these two variables on analytic concepts will be examined later in the monograph.

STUDY 5

The data summarized revealed a moderate association among analytic concepts on CST, reflection over alternative-solution hypotheses, and visual analysis. But the relation between visual analysis and reflection had never been assessed. Since DRT and CST had been administered to groups of second and third graders 1 year earlier (see Study 3), it was decided to retest this group in an attempt to ascertain the relation among all three variables.

The two major components of the new battery were new measures of visual analysis and reflection. The stimuli in the new visual analysis test contained three component parts (in contrast to the stimuli of Study 4 that contained only two components). The background component was a repetitive pattern; whereas, the figural component referred to the form into which the individual-element components fell. That is, the elements were discrete geometric forms that traced out a definite figural form. The figural component, when presented alone, contained no clue as to the original elements that composed it, and the element, when presented alone, contained no clue as to the figural form it assumed in the original stimulus. Figure 7 illustrates one of the items; the stimulus in the upper-left corner was the original design.

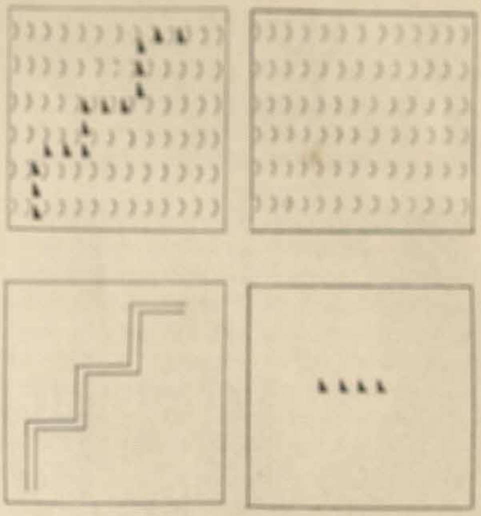


FIG. 7. Sample item from visual-analysis task.

The new task of reflection was similar to DRT, but differed from it in two ways. First, the stimuli were familiar objects (telephone, boat, cowboy) rather than geometric designs. It will be recalled that there was a low-positive relation between accurate performance on DRT and IQ score. It was suggested earlier that this relation might be lower if the items illustrated familiar objects rather than geometric designs. Second, DRT included an immediate memory component, for *S* had to retain the image of the standard for 15 seconds before being allowed to select the variant that he thought was identical to the standard. The new test of reflection did not require memory. The standard and variants were presented simultaneously, and *S* had to select the variant that was identical to the standard. Since memory was not involved the job of detection was made more difficult and the variants were more similar to each other than was the case in DRT. This task was called Matching Familiar Figures (MFF). Figure 8 illustrates two test items.

Procedure

Subjects. The *S*s were 60 boys and 53 girls in Grades 3 and 4 all of whom had been seen a year earlier in Study 3. Each *S* was seen individually at the institute by a female *E*.

Administration of Tests.

1. Visual analysis: Task 1—The *S* was required to learn four different nonsense syllables to each of four designs. The *S* was then presented with cards containing the four nonsense syllables and the stimuli that illustrated the separate components. The transfer task contained two examples each of the ground, figural, and element components (24 stimuli in all)

PSYCHOLOGICAL
LIBRARY
CALCUTTA-19

Bureau of Edcl. & Psyl. Research
Date 9.3.81
J 742

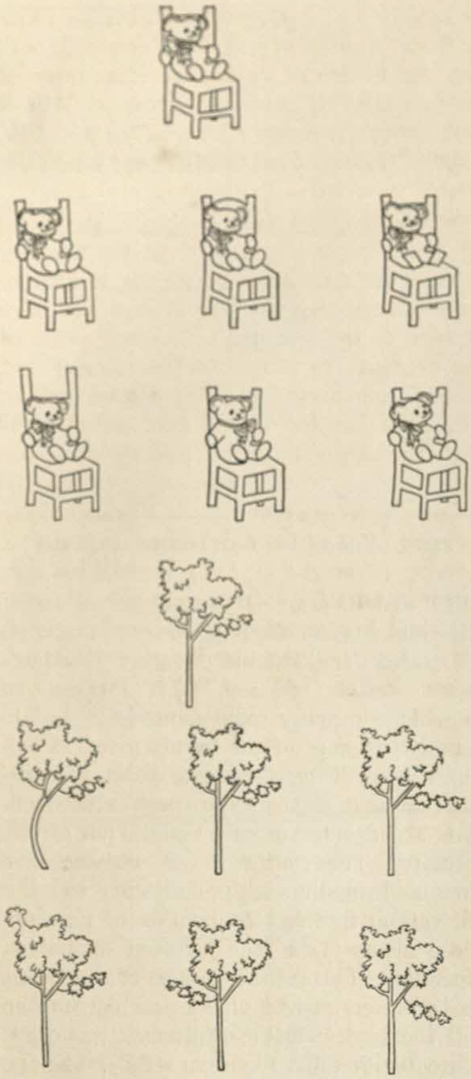


FIG. 8. Sample items from MFF.

arranged in a random order. The *S* was not told if his answers to the separate components were correct.

2. Visual analysis: Task 2—After an intermediate task (MFF) *S* was asked to learn a new set of nonsense syllables to four new stimuli. The procedure was the same as for Task 1. After reaching criterion *S* was shown the separate components (24 in all) and asked to apply the correct nonsense syllable.

The variables scored for each of the tasks were trials to attain criterion; number of ground, figural, or element components correctly labeled on the transfer task; and mean response time for the delay between presentation of the individual ground, figural, or element components in the transfer task and *S*'s answer. It was hoped that response time in the transfer situation would be related to decision times in other situations.

3. MFF—This test was administered between the two visual analysis tasks. The *S* was asked to select the one stimulus from the array of six variants that was identical to the standard. The *S* was given two practice items followed by 12 test items. If *S*'s initial selection was incorrect, he was told he was wrong and was asked to try again. The variables scored were response time to first selection, total errors, and number of test items correct on the first choice.

4. Vocabulary—Each *S* was given the vocabulary scale of WISC. Several months after this session each *S* was also administered the information and similarities scales of WISC and average scaled scores for the three tests were computed.

Results

Visual-Analysis Scores. The figural aspect of the stimulus was clearly the preferred focus of attention (see Footnote 7). All *S*s were most accurate in labeling this component during the transfer trial for Task I. Figural components were also the best recognized for Task 2, but the differences were not as marked. For Task 1 the difference between correct recognition of figure versus element was significant ($p < .01$) as was the comparison between figure and background ($p < .01$, by t tests). For Task 2 these differences were in the same direction but reduced in magnitude.

As might be expected, many *S*s realized after Task 1 that they were not attending to the ground and element components and correct recognition of these categories increased from Task 1 to Task 2, while correct figure recognition decreased slightly. Most of the children shifted their preferred focus of attention from figure to ground and element components after Task 1.

Sex Differences. The boys were more likely than the girls to analyze the stimulus during learning; the differences were most marked for figural and element components for Task 1. The sum of correct background, figure, and element recognitions for each task separately or for both tasks was higher for boys than for girls ($p < .01$ for each comparison). This result matches the sex difference found in the earlier study with simpler stimuli and supports the general belief held by psychologists that school-age boys are more analytic than girls.

There was no sex difference in errors or average response time on MFF.

Age Differences. As expected, the fourth graders had slightly better visual analysis scores on both tasks ($p < .01$) and made less errors on MFF ($p < .01$) than the third-grade children.

Intercorrelations among the Variables. A matrix of correlations among the variables derived from MFF and visual analysis was computed for sex and grade separately. The direction and general magnitude of these correlations were similar for boys and girls in the two grades (see Footnote 7). Trials to criterion on the learning phase of the visual-analysis tasks were unrelated to recognition of the components or to MFF scores, and number correct on MFF was highly related to total errors. Trials to criterion and number correct on first attempt on MFF were eliminated from further statistical analyses.

In order to assess the relation between reflection and visual analysis, and the generality of the early measure of reflection (DRT) to the more recent measure (MFF), intercorrelations were computed among the major variables derived from Study 3 and the present study. Complete data on all tests were available for 30 younger boys and 30 older boys; 28 younger girls and 25 older girls. The 16 variables were derived from CST, DRT, MFF, and visual analysis tests and the mean scaled score on the information, vocabulary, and similarities subtests of WISC. The latter measure was an index of verbal skills. The matrix of coefficients appears in Table 7 for each of the four groups.

Stability of a Reflective Attitude. The two DRT response time measures were each moderately predictive of response time on MFF 1 year later (average $r = .44$, $p < .001$ with DRT I; $r = .34$, $p < .001$ with DRT II). The error scores on DRT I and II were less consistently related to MFF errors than the response times, and the generality was best for the younger boys and the older girls. The average correlations between major errors on DRT I or II and MFF errors were .28 ($p < .01$) and .40 ($p < .001$). As expected, there was a high-negative relation between MFF errors and response time (r's $= -.66$ and $-.65$ for boys; r's $= -.47, -.60$ for girls). The remarkable consistency of a reflective disposition was evidenced by the fact that response time on DRT II predicted errors on MFF 1 year later (average $r = -.29$; $p < .01$). Finally, response time to MFF was minimally related to verbal skills (average $r = .11$).

Significance of Visual Analysis. The scores on the first transfer task were the primary index of the spontaneous tendency to analyze the geometric designs. For after *S* had been exposed to the first test he became aware of the necessity to attend to the element and ground components of the original design. The scores on the second task reflected, in part, the ability to profit from earlier learning.

Correct recognition of the element components on Task 1 was related to analytic concepts (average $r = .23$; $p < .05$), low major errors on DRT I (average $r = -.25$; $p < .01$), and low errors on MFF (average $r = -.23$; $p < .05$), but was independent of response time on DRT I or MFF (average $r = .01, .13$) and verbal skills (average $r = .09$).

Reflection was independent of visual analysis, yet each of these variables contributed variance to the more complex products of analytic concepts and recognition errors. A multiple correlation of .56 obtained for older girls when analytic concepts was the criterion and correct recognition of elements and response time to CST were individual predictors. This coefficient is to be contrasted with the correlations of .45 and .39 for the individual predictors with analytic concepts. When major error score on DRT I was the criterion and response time and recognition of elements were the predictors, the multiple coefficient was .52 for younger boys and .62 for older girls. When MFF error score was the criterion, the multiple coefficient was .73 for younger boys. Thus, response time and visual analysis each contributed unique variance to recognition errors and analytic concepts. Each of these variables was relatively independent of verbal ability, suggesting that they are basic cognitive dispositions.

It is of more than passing interest to note that response time on the transfer series of the visual-analysis task (the time *S* took to

TABLE 7

INTERCORRELATIONS AMONG CONCEPTUAL VARIABLES

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Younger subjects																
1. CST, Analytic concepts	—	−.26	−.17	−.05	−.28	−.13	.10	−.28	−.04	−.23	.01	.44	.02	.47	.51	.23
2. DRT I, Major errors	−.52	—	.78	−.39	.60	.67	−.38	.59	−.50	−.08	−.36	−.35	−.40	−.30	−.42	−.28
3. DRT I, Minor errors	−.36	.78	—	−.29	.52	.55	−.25	.45	−.42	−.15	−.21	−.21	−.31	−.07	−.27	−.19
4. DRT I, Response time	.47	−.43	−.46	—	−.41	−.35	.75	−.15	.26	−.01	.13	.05	−.10	.20	−.01	.03
5. DRT II, Major errors	−.55	.56	.31	−.33	—	.50	−.41	.48	−.37	.14	−.15	−.19	−.06	−.06	−.32	−.11
6. DRT II, Minor errors	−.47	.43	.28	−.21	.64	—	−.38	.63	−.65	−.06	−.20	−.13	−.35	−.37	−.53	−.27
7. DRT II, Response time	.37	−.36	−.37	.54	−.24	−.27	—	.08	−.01	.04	.01	−.08	.11	.16	.19	−.11
8. MFF, Errors	−.21	.09	−.09	.00	.53	.52	−.23	—	−.66	.08	−.28	−.39	−.25	−.33	−.28	−.53
9. MFF, Response time	.29	−.27	−.21	.48	−.23	−.15	.44	−.47	—	−.10	.29	.17	.21	.15	.22	.36
10. Visual Analysis I, Ground correct	−.24	−.25	−.25	−.11	−.07	.15	−.22	.26	−.15	—	.30	.13	.35	.15	.31	−.39
11. Visual Analysis I, Figure correct	−.02	−.15	−.04	.02	−.21	−.02	.05	−.19	.45	.29	—	.49	.43	.35	.42	−.14
12. Visual Analysis I, Element correct	−.13	.02	−.03	−.09	.16	−.11	−.34	−.13	−.13	.13	−.35	—	.37	.47	.49	.16
13. Visual Analysis II, Ground correct	.20	−.21	−.18	−.02	−.48	−.48	.33	−.44	.30	.01	.32	−.10	—	.25	.52	.23
14. Visual Analysis II, Figure correct	.09	−.09	−.04	−.24	−.18	−.40	−.18	−.36	.14	−.15	.41	.01	.39	—	.49	.25
15. Visual Analysis II, Element correct	.14	.04	.03	−.02	.03	−.09	−.11	−.21	.25	.07	.17	−.05	.21	.20	—	.02
16. WISC, Verbal skills	.34	−.44	−.43	.14	−.44	−.32	.14	−.28	.22	.14	.14	.09	.42	.22	−.13	—
Older subjects																
1. CST, Analytic concepts	—	−.44	.02	.32	−.36	.01	.16	−.04	−.05	.16	.08	.10	.03	.23	.08	.30
2. DRT I, Major errors	−.32	—	.08	−.11	.44	−.02	.01	−.02	.31	−.26	.01	−.21	−.17	−.07	−.05	−.38
3. DRT I, Minor errors	−.29	.68	—	−.21	.41	.43	−.07	.08	.09	.12	.02	.06	−.12	.09	−.17	−.21
4. DRT I, Response time	.16	−.51	−.49	—	−.14	.00	.51	−.41	.42	−.23	−.08	−.07	−.43	−.12	−.18	.08
5. DRT II, Major errors	−.37	.84	.64	−.51	—	.32	−.16	04	.09	−.10	−.23	−.27	−.16	−.19	.01	−.18
6. DRT II, Minor errors	−.11	.65	.59	−.48	.68	—	−.35	.01	.05	−.09	−.16	.12	.03	−.12	−.08	24
7. DRT II, Response time	.34	−.34	−.47	.71	−.31	−.24	—	−.42	.41	−.22	−.17	−.04	−.25	−.13	−.13	−.40
8. MFF, Errors	−.18	.35	.59	−.43	.50	.38	−.49	—	−.65	.05	−.06	−.25	−.01	−.20	.21	.10
9. MFF, Response time	.30	−.33	−.34	.56	−.37	−.48	.45	−.60	—	−.09	−.12	.09	.08	.00	−.15	−.13
10. Visual Analysis I, Ground Correct	.04	−.21	−.07	.14	−.28	−.20	−.02	−.06	.05	—	.59	.44	.55	.28	.25	.09
11. Visual Analysis I, Figure correct	.29	−.28	−.18	−.03	−.52	−.21	−.14	−.14	.13	.37	—	.53	.24	.50	.25	.05
12. Visual Analysis I, Element correct	.45	−.41	−.11	.12	−.49	−.49	.18	−.11	.36	.00	.30	—	.33	.49	.05	−.06
13. Visual Analysis II, Ground correct	−.14	−.38	−.15	.40	−.35	−.44	.17	−.43	.46	.03	.16	.40	—	.13	.26	.18
14. Visual Analysis II, Figure correct	−.40	−.05	−.01	.12	−.03	−.27	−.05	−.10	.08	.19	.15	−.10	.32	—	−.13	.16
15. Visual Analysis II, Element correct	.29	−.23	−.20	.38	−.34	−.31	.57	−.36	.50	−.17	−.03	.20	−.03	.05	—	.33
16. WISC, Verbal skills	.06	−.47	−.37	.07	−.56	−.29	−.06	−.40	−.03	.21	.30	.17	.16	.01	.02	—

Note.—In the upper portion of the table younger boys ($N = 30$, $r = .35$, $p < .05$, two tailed) are to the right and above the diagonal while younger girls ($N = 28$, $r = 37$, $p < .05$, two tailed) are to the left and below the diagonal.

In the lower portion of the table older boys ($N = 30$, $r = .35$, $p < .05$, two tailed) are to the right and above the diagonal while older girls ($N = 25$, $r = .39$, $p < .05$, two tailed) are to the left and below the diagonal.

decide what nonsense syllable to apply to the components of the design) was positively related to response time on MFF ($r = .68$ for Task 1, and $r = .33$ for Task 2). A conceptual tempo variable—the tendency to delay before answering—operated not only in situations in which the alternatives were all visually present (as in MFF), but also in problems in which *S* had to select one of four responses generated from imagery. The common process in these two task situations was the selection of one response from a set of simultaneously available alternatives.

STUDY 6

Previous studies have presented convincing evidence for the pervasiveness and significance of a decision-time variable. To postulate a psychological dimension carries a responsibility for inquiring into the possible bases for differences in this predisposition. One variable that might be relevant is anxiety generated by the child's expectation of rejection or disapproval from *E*. The effect of experimenter rapport tends to be minimal when a task is moderately difficult and interesting and maximal when a task is too easy. Tasks like DRT or MFF are difficult, but not frustrating, and seem to hold the interest of the children. Thus, experimenter rapport should have a minimal effect on performance on these tests. It seemed appropriate, nevertheless, to investigate the possible effect of this variable. For it could be argued that apprehension over acceptance by *E* would lead most children to modify their behavior in order to reduce this particular source of anxiety. There are at least two kinds of behaviors that might serve this defensive function. The child might respond more quickly because he was less able to tolerate the period of silence between the presentation of the problem and his production of a solution. An alternative argument assumes that anxiety over rejection would increase the child's need to obtain *E*'s approval, and to make potential failure more unpleasant than it ordinarily would be. The child might become more reflective under these conditions in order to insure success.

It is reasonable, therefore, to argue that *E*'s acceptance might lead either to increased or decreased reflection. This study assessed the differential effect of impersonal versus emotionally reassuring testing conditions on instruments tapping an impulsive versus a reflective attitude.

Procedure

The *Ss* were 60 boys and 60 girls in the third grade of a public school. Children were paired on sex and IQ and one member of the pair was assigned to the "impersonal" testing group; the other to the "reassuring" testing group. One male *E* was used for all subjects. Under the impersonal testing conditions designed to create anxiety over experimenter acceptance *S* was told to sit quietly at a table for a little while. The *E* then worked at another table for 2 minutes and did not interact with the child. The *E* neither smiled nor talked to *S* during the test administrations except to tell him whether he was correct or incorrect on a test item. Social interaction was minimal and *E* displayed no facial or verbal signs of warmth or acceptance.

Under the reassuring testing conditions *E* spent the first 2 minutes chatting with *S* about his hobbies and interests, and tried to establish a friendly relationship with him. The *E* smiled often during the testing administrations and warmly praised *S* when the latter was correct.

Administration of Tests.

1. DRT—The DRT was administered under conditions described earlier.

2. Story Recall—The *E* read a paragraph of 106 words to *S* with the prior instruction that *S* would have to repeat the paragraph. The paragraph contained several incongruous elements and specific details. The purpose of this task was to assess the differential effect of the two testing conditions on immediate recall. The story was constructed to contain 15 critical-meaning units, and the child's score was the number of units recalled.

3. MFF—The MFF was administered under conditions described earlier.

4. Haptic visual matching (HVM)—This new test was designed to measure the impulsivity-reflection dimension with a cross-modal task—the haptic and visual modes. The child first explored tactually a three-dimensional wooden form to which he had no visual access. After this exploration *S* selected from a visual array containing five similar stimuli the one pattern that was an identical representation of the form he explored haptically. This task was devised in order to inquire whether the dimension of impulsivity derived from completely visual tasks (DRT or MFF) would generalize to a situation in which the initial comprehension of the stimulus was obtained through a different modality. A second rationale was that this test furnished an objective index of a scanning time. For the time *S* took to explore the wooden form could be regarded as a measure of the amount of time taken to comprehend

the new stimulus. It was hoped that this task would furnish information on the following question: Is the occurrence of recognition errors due primarily to fast response times in the final selection process, or does amount of scanning of the original stimulus contribute some variance to error scores?

The task contained 20 stimuli, 10 of which were geometric forms; the remaining 10 were objects with which the child was familiar (tree, child, hand, airplane, automobile, pipe, scissors, shirt, cat, pig). The child was told to palpate the wooden form with his fingers as long as he wished and to withdraw his hand from in front of a black curtain when he was finished. He was then shown a card illustrating five "ink-line" drawings and asked to select the visual form that corresponded to the one he had palpated. The *E* recorded palpation time of the standard, response time to the first selection, and number of errors for each item. Figure 9 illustrates the visual arrays for two items.

These four tests furnished 18 variables for analysis. Intelligence might have interacted with the two testing conditions, and each of the four groups (two sexes and two testing conditions) was divided into high- and low-intelligence groups based on scores from CTMM which had been gathered by the school several months earlier. The two intelligence groups should be regarded as normal versus bright normal, for only two children had IQ scores below 90.

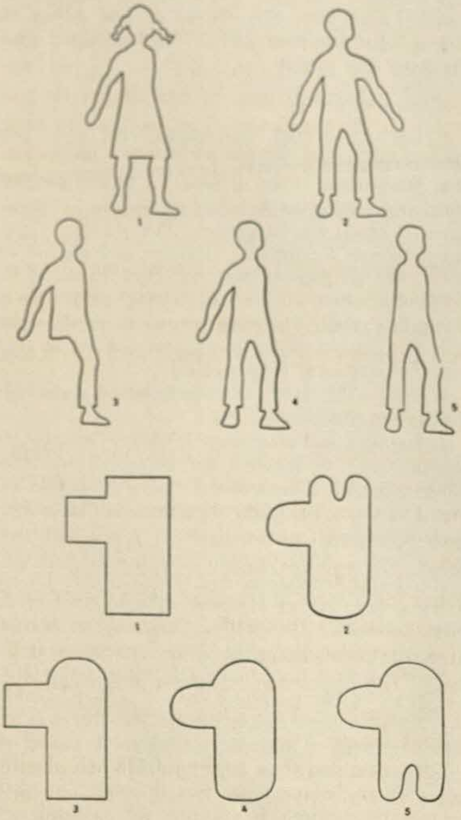


Fig. 9. Sample items from HVM.

Results

Analyses of variance were performed for each of the variables, and the main results are summarized below (see Footnote 7).

Effect of Testing Conditions. There was no significant effect of impersonal versus reassuring testing conditions on any of the three error variables. Moreover, the only response-time variable for which testing conditions had a significant simple effect was response time on HVM, with impersonal conditions producing longer response times. But this effect did not hold for response times to MFF or DRT. Finally, testing conditions had no effect on story recall. In sum, testing conditions had a minimal effect on both response time and error scores, supporting the earlier statement that warm or cold rapport is not particularly relevant for the reflection dimension when the test situations are moderately challenging and of interest to the child.

Intercorrelations among the Variables. In order to assess the consistency of response time and errors for the three tasks, intercorrelation matrices were computed for each of the four groups with 30 *S*s in each group (i.e., two sexes by two testing conditions). The direction and magnitude of the correlations were remarkably similar for both testing conditions, and average *r*'s (z' transformations) were computed for *S*s in the two testing conditions with sexes kept separate. The matrix of coefficients for the nine major variables appears in Table 8.

These data, like those presented earlier, present a picture of consistency for both response time and errors across the varied tasks. Errors on DRT, MFF, and HVM were positively correlated for both sexes (r ranged from .33 to .52, all significant at $p < .01$). Similarly, response times to DRT, MFF, and HVM were highly intercorrelated (r ranged from .48 to .82), the correlations being higher for females than for males. As usual, the intertask consistency for response times was greater than the consistency for errors.

Response times were inversely correlated with error scores for all tests, but the relation

TABLE 8

INTERCORRELATIONS AMONG CONCEPTUAL VARIABLES
(Average *r* for two testing conditions)

Variable	1	2	3	4	5	6	7	8	9
1. DRT, Errors	—	.53	.33	−.28	−.33	−.35	−.39	−.13	−.30
2. MFF, Errors	.49	—	.52	−.22	−.57	−.51	−.35	−.35	−.47
3. HVM, Errors	.44	.48	—	.07	−.15	−.28	−.50	−.37	−.37
4. DRT, Response time	−.33	−.34	−.25	—	.48	.57	.18	−.22	.01
5. MFF, Response time	−.32	−.51	−.43	.79	—	.59	.32	.07	.05
6. HVM, Response time	−.38	−.45	−.47	.75	.82	—	.56	−.09	.14
7. HVM, Palpation time	−.35	−.42	−.40	.22	.43	.55	—	.24	.08
8. Story Recall	−.14	−.26	−.30	.00	.10	.05	.14	—	.29
9. IQ (CTMM)	−.25	−.21	−.35	−.01	.15	.03	.00	.45	—

Note.—Boys are to the right and above the diagonal ($N = 60$) and girls are to the left and below diagonal ($N = 60$). $r = .25$ for $p < .05$, two tailed; $r = .33$ for $p < .01$, two tailed; $r = .41$ for $p < .001$, two tailed.

was best for MFF. Additional data from several other samples also revealed that the negative relation between response time and errors was better for MFF than for DRT or HVM. The MFF task is more difficult than the other two tests and this is one reason for this relationship. Once again, the relation between errors and response time held across tasks. For DRT errors were as highly associated with response time on HVM (*r*'s= −.35 and −.38) as they were with response time on the DRT itself (*r*'s= −.28 and −.33), indicating the pervasiveness of a decision time variable.

Palpation time on HVM, the time *S* took to explore the wooden form, was highly related to HVM response time (*r*'s = .56 and .55). Children who responded quickly took less time to explore the standard. Moreover, palpation time predicted DRT errors (*r*'s = −.39 and −.35) and MFF errors (*r*'s = −.35 and −.42) and, for boys, was a better predictor of HVM errors than HVM response time (*r* = −.50 versus −.28).

Relationship to IQ. There were low-negative correlations between IQ score and errors on each of the three tasks (*r* ranged from −.21 to −.47), but no relationship between response time and IQ (*r* ranged from −.01 to .15, all nonsignificant). Thus, recognition errors were jointly influenced by the multiple processes implied by high IQ and by decision time. Some of the basic variables implied by high IQ score are special problem-solving skills, rich vocabulary resources, confidence in one's ability to solve problems, and the motivation to do well on intellectual tasks. The multiple correlations with DRT, MFF, or HVM errors as criterion, and response time and IQ as individual predictors were high. The multiple coefficient for MFF errors was .72 for boys and .54 for girls; for DRT it was .41 for boys or girls. The multiple correlation for HVM was slightly higher for boys if palpation time, rather than response time, was the predictor. The coefficients were .60 for boys and .53 for girls when palpation time was used; the coefficients were .46 for boys and .58 for girls when response time was used. When intelligence and response time were jointly taken into account it was possible to account for approximately twice as much variance in error scores as when only one of these variables was used as a predictor.

Discussion

The minimal effect of impersonal versus reassuring testing conditions upon response time and recognition errors suggests that these predispositions are not easily changed by experimenter rapport and are fundamental response tendencies in the child. The consistency of recognition errors and decision times across three different tasks is another demonstration of the pervasiveness of an individual difference variable which might be called "conceptual tempo."

The palpation time scores for HVM indicate that *S*s who explore the stimulus rapidly are also apt to respond quickly and are likely to make more errors. It would be important to determine if visual scanning times on MFF or DRT, the analogue of palpation time, manifest the same relationship to response time and errors as palpation time does on HVM.

STUDY 7

The negative relation between analytic concepts and impulsive offering of incorrect solutions is further supported by data from a small group of Fels children.

Procedure

Subjects. The *Ss* were 34 boys and 22 girls from the Fels Research Institute's population (age range from 7-1 to 11-4). Each *S* was seen individually by a female *E* [11] for two 1-hour sessions and administered the following test battery during the first session.

Administration of Tests.

1. CST—Each child was administered CST and the variable scored was the number of analytic responses.
2. Embedded Figures Test (EFT)—Each child was administered 10 items from the Witkin Embedded Figures Test using standard instructions (Witkin et al., 1962). The major variables scored were (*a*) total time to solution and (*b*) the total number of incorrect solution hypotheses.

The tests administered during the second session were as follows.

3. Word-association test—Each *S* was administered the 40-item word-association test described in Study 1. The major variables scored were frequency of coordinate nouns, coordinate verbs, and antonyms to adjective stimuli.
4. Figure-ground visual-analysis task—Each *S* was administered a visual-analysis task similar to that described in Study 4 in which only two components with figure and repetitive background were used, but five rather than four geometric designs were presented.
5. PDT—Each *S* was administered PDT as described in Study 4. The *S* was allowed to continue responding to each of the 30 items until he located the correct difference, or for a maximum of 2 minutes. The major variable was the total number of incorrect hypotheses offered.
6. Transfer Trial 2: Visual-analysis task—Each *S* was again presented with each of 10 figure or ground variants derived from the original geometric designs.

Results

A total of 11 variables was derived from these five tests, and these scores were intercorrelated (see Footnote 7). The major relationships can be summarized briefly. For boys analytic concepts were negatively related to number of incorrect hypotheses on PDT ($r = -.32, p < .10$), and EFT ($r = -.49, p < .01$). These data are in essential agreement with the results of earlier studies in supporting the inverse association between analytic concepts and the production of incorrect responses prior to solution. It is important to note that analytic concepts were not highly related to the ability to solve EFT ($r = -.27$ with solution time, not significant). An analytic attitude on CST was more closely related to a strategy of problem solving (i.e., reflection on the adequacy of alternative answers) than to the ability to solve this kind of problem (i.e., detect the disguised stimulus). Since Witkin and his colleagues regard solution time on EFT as the primary index of field independence, the present construct of reflection differs in meaning from Witkin's concept of field independence.[12]

There was a moderately positive relation between analytic concepts and correct labeling of figural elements on the transfer trials (r's = .23 and .28), and this finding supports earlier results. The two error scores (i.e., number of incorrect responses on PDT and EFT) were moderately related to each other ($r = .39, p < .05$), and each had similar correlation patterns with the other variables. For boys EFT errors were negatively related to correct recognition of the figural elements (r's = $-.52$ and $-.47$ for Transfer Trials 1 and 2, $p < .01$).

The pattern of relationships for girls was somewhat different than that obtained for boys. Analytic concepts were not highly related to errors on PDT or EFT, but were positively related to recognition of both figure and ground components for transfer Trial 2 ($r = .39, p < .10; r = .52; p < .05$). As with boys, errors on PDT and EFT were positively correlated ($r = .51, p < .05$), and PDT errors were negatively related to correct recognition of the figural components. Moreover, errors on PDT and EFT were each associated with infrequent production of coordinate nouns, verbs, or antonyms. Thus, an impulsive approach to PDT and EFT was associated with syntactic immaturity for the girls, as infrequent analytic concepts on CST were associated with syntactic immaturity for the first and second grade boys in Study 1.

---

[11] The authors are indebted to Carolyn Cermele and Alice Rabson who administered these tests.

[12] In an unpublished investigation of 22 women between 50 and 60 years of age, we found no relation between accurate recognition of the figural component in the visual-analysis task and solution time on EFT, but a significant negative relation between recognition of the figural component and number of incorrect solutions on EFT ($p < .05$).

The results are in essential agreement with earlier work and strengthen the basis for postulating the two dimensions of reflection-impulsivity and visual analysis.

STUDY 8

Observation of reflective and impulsive boys during the testing sessions suggested that the latter group typically displayed more restlessness and greater distractibility. The sections that follow describe three attempts to obtain more systematic data on the relation between motor restlessness or distractibility and conceptual impulsivity.

Part A

Subjects. The *S*s represented extremes on the distribution of analytic concepts and recognition errors on DRT. The *S*'s were selected from the 72 boys and 63 girls seen in Study 3 who had been administered CST and DRT. Since analytic concepts were related to low error scores on DRT, two groups were composed. The *S*'s in the low-impulsive group were above the median on the distribution of analytic concepts and below the median on the distribution of total errors on DRT I. The *S*s in the impulsive group were below the median on analytic concepts and above the median on DRT errors. This dichotomy yielded 21 boys and 13 girls classified as nonimpulsive and 22 boys and 20 girls classified as impulsive.

Procedure. Each child was observed in his classroom from 8:45 until 10:45 each morning. The observer coded *S*'s behavior only during those times when *S* was seated at his desk and the classroom situation called for attention to a school task. The observer checked the occurrence of each of five behavioral categories during successive 10-second periods. Only one check was permitted for each variable during any 10-second period, but it was possible for all five variables to be checked during a single period. Approximately 60 minutes of coded observational data were obtained on each *S* during each 2-hour session, and 70 of the 76 *S*s were observed on two separate occasions yielding 2 hours of observational data. The scores consisted of the average values for the two sessions combined for the following variables: (*a*) leg movements, (*b*) arm movements, (*c*) trunk movements, (*d*) attention to task, and (*e*) distractibility (i.e., child looks away from the task). The interrater reliabilities for these categories were obtained by having two observers independently code 72 separate 10-minute sessions. The reliabilities (product moment) for each of the five categories ranged from .84 to .96 with a median reliability coefficient of .94.

Results. The variable of primary interest was the ratio of time spent attending to a task divided by time spent distracted. This ratio ranged from .6 to 14.9 (median 3.0) for boys, and from .7 to 3.7 (median 1.9) for girls. The stability of this ratio from Observational Session 1 to Session 2 was moderate ($r = .34$ for boys and .68 for girls, p's $< .05$ and $< .001$). The ratio of task attention to distractibility discriminated the impulsive and nonimpulsive children. The mean for the nonimpulsive boys was 3.9; the mean for the impulsive boys was 2.7 ($p < .05$). For the girls the corresponding mean scores were 2.2 and 1.7 ($p = .06$). The number of limb or trunk movements while seated at a desk did not discriminate the two groups.

Extreme scores on this ratio were excellent predictors of the impulsivity dimension for the boys. Ten boys had ratios of 4.0 or higher on both observational sessions and eight *S*s in this group were nonimpulsive on the tests. It should be noted that the ratio of task attention to distraction was not as highly related to response time on DRT as it was to error and analytic concept scores. Thus, the derived variables of analytic concepts and recognition errors were better predictors of the classroom behavior than the decision time variable.

Impulsive *S*s were likely to display momentary lapses of attention during involvement in a school task; they were more likely to look out the window, gaze at a peer, or orient to a sound during a period when they were working at an academic task. These behavioral observations are in complete agreement with the problem-solving behavior displayed by the children in the laboratory setting. The reflective children seemed capable of greater concentration; they showed less distractibility during the conceptual tasks.

Additional observations of the free-play behavior of the boys supported these classroom observations. Each of the 43 boys in the above study (each was extreme on the impulsivity dimension) was observed for 10 minutes during a school recess period on the playground. Each time the boy initiated gross motor activity (ran, jumped, played on gross motor apparatus) a stopwatch was started. When *S* stopped the activity the watch was stopped. The total time recorded at the end of 10 minutes was an index of the time spent in gross motor activity. There were no differences in activity between second and third grade boys and the data from both grades were combined. There was a significant difference between the two groups with the impulsive *Ss* spending more time in gross motor activity than the nonimpulsives (1.8 minutes versus 3.2 minutes, $p < .05$). These results agree with earlier findings indicating more frequent gross motor behavior by impulsive boys (Kagan et al., 1963).

Part B

Behavioral observations on the sample of Fels boys described in Study 7 support the association between spontaneous displays of task unrelated gross motor behavior and an impulsive attitude.

Procedure. Twenty-seven of the 34 boys between 7 and 11 years of age who were given CST in Study 7 had been observed extensively from birth to age 8 in the child's home and in the Fels experimental nursery school. Most *Ss* were observed for a half day semiannually in the home, and a description of their behavior was prepared after the visit. Beginning at 3 years of age each child was brought to the institute once or twice a year for 2 weeks during which he participated in a small play group of 6–9 children of his own age. The *S*'s behavior was observed systematically, and a lengthy description of his behavior was prepared. These observational protocols were divided into two sets covering birth to 4 years of age and 4–8 years of age. A senior psychology student was trained to rate these protocols for four variables:[13] (*a*) gross motor activity, (*b*) time alone involved in task mastery, (*c*) retaliation to attack by same-sex peers, and (*d*) instigation of physical aggression to same-sex peers. The student made two sets of ratings (on a 7-point scale) for these four variables. One set of ratings was made for age 0–4 and the other for age 4–8. The rater had no knowledge of *S*'s performance on the conceptual tasks while she was making her ratings. No scores were available for 10% of the ratings because of insufficient data. Independent ratings of this material by one of us (JK) indicated satisfactory interrater reliability (r ranged from .85 to .93 for the four variables).

The number of analytic concepts on CST was available for each child, and each *S* was assigned to a high- or low-analytic group depending on whether his score fell above or below the median for his age group. This dichotomy yielded 13 high-analytic and 14 low-analytic boys. The mean scores for the two groups on each of the four behavioral variables for ages 0–4 and 4–8 were compared.

Results. The results are in complete accord with those found on the sample of second and third grade boys, and with data reported earlier (Kagan et al., 1963). Boys high in analytic concepts, in contrast to those with few concepts, displayed less spontaneous gross motor behavior during age 4–8 and more solitary task oriented behavior during both ages 0–4 and 4–8 (p's $< .10$, $< .05$, $< .05$). There was no association, however, between instigation of aggression or retaliation and analytic concepts. The fact that ratings of solitary involvement in task mastery during the first 4 years of life predicted analytic concepts 4 years later suggests that some of the critical antecedents for this conceptual category are present early in development.

An earlier and more extensive investigation of behavioral continuities among Fels subjects (Kagan & Moss, 1962) reported that ratings of hyperkinesis during the period 3–6 years of age predicted phenotypically similar behavior during ages 6–10 and 10–14. Moreover, hyperkinesis during age 6–10 was inversely correlated with ratings of involvement in solitary intellectual mastery among adult men ($r = -.37$, $p < .05$). The fragile empirical link between behavior during the first 5 years and an analytic attitude during the school years suggests that constitutional variables and/or early learning influence this class of conceptual responses.

---

[13] The authors are indebted to Eva Manoff for ratings of these observational data.

Part C

A final set of data based on respiratory patterns to episodes calling for attention corroborates the inverse relation between restlessness and distractibility, on the one hand, and an analytic attitude on the other.

Procedure. Fifty-five of the 62 boys in Study 1 were seen for a 45-minute experiment during which heart rate and respiration were monitored for 16 1-minute episodes. Four of the episodes required *S* to attend to a simple visual or auditory stimulus; 4 required mental elaboration; with 8 1-minute rest periods interposed between each of these work episodes. A complete description of the procedure appears in another report (Rosman, 1962). A brief summary of the procedure follows.

Each *S* sat beside a female *E* (BLR) who administered the following procedures:

1. Rest (RST).
2. Color wheel (COL)—The *S* was asked to watch a red sector in a color wheel while the entire wheel revolved for a 90-second period.
3. Rest (RST).
4. Make-up-stories (STY)—The *S* was given a picture for 10 seconds after which it was removed, and *S* was asked to think up a story silently during the remaining 20 seconds. This procedure was repeated three times for a total of 90 seconds.
5. Rest (RST).
6. Word association (WRD)—The *S* was read a word or short phrase and asked to think up a picture silently to the word or phrase.
7. Rest (RST).
8. Stick-on-floor (STK)—The *S* listened to the sound of a stick hitting the floor at irregular intervals.
9. Rest (RST).
10. Make-up-story (STY)—The *S* was given three different pictures as in Number 4 above.
11. Rest (RST).
12. Listen to noise (NOI)—The *S* was asked to attend to increases in the intensity of the masking noise that was present in the room. The increases were brief and irregularly spaced over a 90 second period.
13. Rest (RST).
14. Attend to light (LGT)—The *S* was asked to watch momentary increases in the intensity of light in the room.
15. Rest (RST).
16. Make-up-story (STY)—The *S* was given three new pictures and asked to make up stories as in Number 4 above.

For each of the 16 episodes, four basic variables were coded: respiratory period, standard deviation of respiratory period, average cardiac period, variability of the cardiac period. The major results were that under conditions of attention (COL, STK, NOI, LGT), in contrast to the contiguous rest periods (RST), respiratory variability was markedly less and cardiac rate markedly lower for the majority of the children (Kagan & Rosman, in press). However, boys high on analytic concepts displayed this general effect even more clearly than less analytic children. Figure 10 shows the average standard deviation of respiratory period for high- and low-analytic boys (based on median split) over the 16 episodes. Each value is based on the first 20 respiratory cycles in each of the 16 episodes.

Results. Respiratory variability was lowest for all *S*s for those situations in which the child had to attend to visual or auditory signals (COL, WRD, STK, NOI, LGT). But analytic boys breathed at a more regular rate for the first eight episodes than the less analytic ones ($p < .01$). The negative relation between analytic concepts and respiratory variability during an attention period was highest for the color wheel (COL; $r = -.38$, $p < .01$). The major causes of increased respiratory variability were restless movements of the trunk, sighing, and lack of attentiveness to the test stimuli. Observer ratings of restlessness and distractibility during the experiment suggested that the analytic boys were less restless and more attentive. The respiratory data furnished more elegant and objective support for these cruder observations. In sum, on the playground, in the classroom, or in a restrictive laboratory situation, analytic boys were less restless, less distractible, and less active than their nonanalytic counterparts.

An unpublished study by Irving Sigel furnishes additional support for these conclusions. Sigel presented pictures to 22 boys and 22 girls of kindergarten age who came from upper-middle-class families. Each of the 20

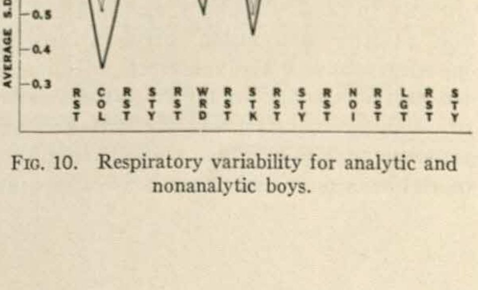


FIG. 10. Respiratory variability for analytic and nonanalytic boys.

pictures displayed a standard and three choice stimuli. The child was asked to select one of the three choice stimuli to "go with" the standard. The children's school behaviors were rated by their teachers. The boys who produced analytic conceptual groupings (i.e., their selections were based on similarity in some differentiated and objective component between the standard and one of the choice stimuli) were rated as more controlled emotionally than the less analytic boys. The correlation between "emotional control" and analytic concepts was .45 ($p < .05$) for boys, but −.20 for girls. These data on a group of upper-middle-class kindergarten children are consonant with those reported here for a more representative group of second and third graders. The lack of a relationship between emotional control and analytic concepts among girls supports previous suggestions that the relationship among impulsivity, restless motor behavior, and analytic concepts is somewhat different for boys than for girls.

DISCUSSION

These studies have been concerned with four related questions: the behavioral significance and basic determinants of analytic concepts, and the generality and stability of reflection impulsivity and visual analysis.

Meaning of Analytic Concepts

Although the empirical work began with investigations of analytic concepts on CST, it has become clear that this variable has multiple antecedents, two of which are reflection and visual analysis. When a set to reflect was created experimentally (Study 2), number of analytic concepts was increased. Moreover, long response times on CST and DRT were positively correlated with analytic responses. The role of visual analysis in facilitating analytic responses was supported directly in Studies 4 and 5, and indirectly in the positive association between analytic groupings and initially correct selection of the correct variant on DRT. When number of analytic concepts was the criterion and response time to CST and correct recognition of elements on visual analysis were separate predictors, the multiple-correlation coefficient reached .56 for one group of fourth grade girls. The overt behavioral correlates of an analytic attitude, among boys, included a capacity for sustained concentration in intellectual tasks, absence of an extreme degree of task irrelevant gross motor activity, and inhibition of impulsive-solution hypotheses in varied problem tasks.

In the light of all the data the following interpretation of the analytic concept is offered. The most obvious way to pair stimuli on most of CST items involves a thematic or functional relationship between two objects. The dominant association to the watch, man, and ruler is a functional linking of the watch and man, or the ruler and man; the dominant association to the house, matches, and pipe is a functional bond between the matches and pipe. If the child is to produce an analytic concept, he must suppress these initially strong, "popular" associations and reflect over alternative responses. If, in addition to a reflective attitude, he also has a predilection for visual analysis, he is likely to produce an analytic concept. Thus, reflection over alternative solutions together with a predisposition to analyze visual arrays are intuitively reasonable determinants of analytic concepts. The empirical data support the intuitive argument. It is to be noted that analytic concepts, visual analysis scores, and indexes of reflection each display a linear increase with age during the school years (Kagan, in press).

Significance of Reflection Impulsivity

A second, and probably more significant, set of findings is contained in the stability and generality of a tendency toward reflection. Long response times on the varied perceptual recognition tasks showed a greater stability over time and more impressive intertask consistency than indexes of an analytic attitude or recognition-error scores. The reliability of response times over a 9-week period was in the high 70's, and there was remarkable generality across tasks as varied as DRT, MFF, and HVM. It is to be noted that the high degree of consistency for response times is limited to this particular kind of problem (i.e., many response alternatives available simultaneously and the correct alternative is not immediately obvious). This consistency should not necessarily occur on tasks with only one alternative or on tasks for which the alternatives are not quickly and simultaneously available. Response times to simple rote questions (e.g., What is your name? What is the month?); to difficult arithmetic questions; or on simple motor-reaction-time tasks should not necessarily correlate highly with each other or with response times on tasks like DRT or MFF.

Reflection is defined semantically as the consideration of alternative solution hypotheses (either classifications or problem-solving sequences) when many alternatives are available simultaneously. Reflection does not refer to delay that is the result of fear of failure, timidity, or inability to generate any solution. The tasks used in this work minimize fear of failure, but further efforts to guarantee this control are necessary.

Significance of Recognition Errors

Recognition errors, like analytic concepts on CST, are multiply determined with reflection as the major variable and visual analysis and verbal skills as secondary factors. Error scores on DRT or MFF were consistently related to verbal ability; whereas, response times were independent of verbal skills. The potpourri of processes implied by the phrase "verbal intelligence" influence quality of performance on problems that do not require complex reasoning operations or rich verbal expression. A verbal intelligence index encompasses a variety of more fundamental factors such as expectancy of task success, availability of verbal labels for unique classification of stimuli, special problem-solving skills and strategies, and motivation to perform well. When a child believes that a problem has a definite answer and requires concentration for success, it is usually the case that a global index of intelligence relates positively to performance score when the sample tested contains an adequate range of IQ scores (e.g., 90–120). The tendency to reflect on alternative solution possibilities is only one problem-solving strategy that has only indirect implications for performance. The important distinction between a reflective strategy and success with a problem was supported by the finding that solution time on EFT, which measures quality of performance, was unrelated to analytic concepts; whereas, generation of incorrect hypotheses on EFT was positively related to an analytic attitude.

Antecedents of Reflection Impulsivity

A tendency toward reflection versus impulsivity displayed a stability over time and a generality across tasks that is unusual for psychological attributes and tempts one to conclude that this disposition is a basic component of a child's behavioral organization. It is critical, therefore, to understand the antecedents of this predisposition. Why are some children more reflective than others? Three possibilities deserve exploration: constitutional predispositions, degree of involvement in tasks, and expectation of failure.

There is growing evidence suggesting that one of the possible consequences of minimal brain damage during the perinatal and early postnatal periods is increased restlessness and distractibility during the preschool and early school years. It is possible that extreme impulsivity at age 6 is a partial consequent of subtle cerebral insult early in life. The positive relation between level of gross motor activity during ages 4–8 and nonanalytic concepts at age 8 suggests that the basic determinants of impulsivity may be present early in development. Biological processes, unrelated to CNS insult, could also account for large individual differences in activity level and, as an epiphenomenon, conceptual impulsivity. Schaefer and Bayley (1963) found that very active 10-month-old boys were rated as low on attentiveness during the period 27–96 months of age. Preliminary results from an intensive longitudinal study of infant activity and attention at the Fels Institute suggest that vigor of motor activity and degree of sustained attention to visual stimuli are inversely correlated, and each is moderately stable from 8 to 56 weeks of age. Moreover, the boys were more active and less attentive than the girls to visual presentations of pictures and blinking light patterns. The evidence is still thin, but strong enough to support the conjecture that some of the basic determinants of conceptual reflection versus impulsivity can be previewed in the activity and attentional behavior of the infant.

A second set of psychological factors that might influence the reflection dimension is degree of involvement in the task. The child who has high standards of performance should be more likely to reflect on alternative hypotheses. The child who cared less about the quality of his product should be likely to accept his first association on CST and adopt an impulsive trial-and-error strategy to DRT or MFF.

A final hypothesis concerns the child's anxiety over his ability to perform well in the tests of reflection impulsivity. Most of the children appeared to have a desire to do well on the tasks presented and to convince *E* that they were able to perform adequately. Let us look closely at the psychology of the situation the moment *E* presents a DRT item to the child. The child who is anxious about his ability and expects to fail may be unable to tolerate the period of silence that is an inevitable concomitant of response selection. He may fear that the silence will be interpreted as an indication of his inability to provide a correct answer immediately, and in order to reduce this tension he may offer an answer impulsively. The child who is minimally anxious about his ability on these tasks will be less anxious over the possibility that *E* will disapprove of him and will be more able to tolerate the delay between presentation of a problem and his reporting of an answer. Delay in the service of reflection may be facilitated when the child expects success, for this expectation makes it easier to tolerate the silence that accompanies consideration of alternative answers. These three hypotheses are not mutually exclusive and all three might contribute variance to the reflective dimension.

Methodological Implications of Reflection and Analysis

A response to a psychological test—be it an ink blot interpretation, a story to a picture, or the selection of an answer in a multiple-choice inventory—usually involves processing of information and the selection of one best response from a set of alternatives. Motives, conflicts, and defenses are not the only factors that affect the final cognitive product in these test situations. The dimensions of reflection and visual analysis influence the manner in which information is classified and, consequently, the content of the final response. The acknowledgement of individual differences in mode-of-information processing leads one to look at personality test data, for example, through lenses that differ from the conventional kind. To illustrate, Fisher and Cleveland (1958) use an *S*'s ink-blot interpretations as an index of a variable called "barrier response" which is assumed to reflect a person's basic conception of his body. Study of the responses that comprise the barrier score reveals that they often refer to well-delineated objects qualified by descriptive adjectives (e.g., woman with a dress on, man with armor). Previous studies (Kagan et al., 1963) and the data summarized in this monograph indicate that analytic children are likely to produce interpretations that are well delineated and contain adjectives describing physical attributes of objects. Nonanalytic children are more likely to give amorphous responses having minimal articulation and not qualified by descriptive adjectives (blood, ice cream, mashed bug). These latter percepts are called "penetration" responses by Fisher and Cleveland. The occurrence of *barrier* and *penetration* responses may be a function, in part, of the degree of analysis of the ink-blot stimulus and degree of reflection over alternative interpretations and only minimally related to body image.

A second methodological implication of this work touches the validity of the assumption that factor analysis of group-administered tests can yield the basic components of intellective activity. It is difficult to obtain sensitive indexes of response time from group-administered test procedures. The data summarized suggest that the reflection-impulsivity dimension may be a basic variable, but it is not likely that a factor analysis of the data derived from most group administered test batteries would yield a "reflection" factor.

Problem Solving

A final implication of the reflection-impulsivity dimension concerns the gradual establishment of permanent attitudes toward problem solving and strategies of solution with novel-problem tasks. In complex problems with alternate routes to solution reflection upon the probable validity of varied solution sequences is critical for the ease with which success is achieved. The child who does not reflect upon the differential validity of several solution possibilities is apt to implement mentally the first idea that occurs to him. This strategy is more likely to end up in failure than one that is characterized by reflection. For the impulsive child who reaches a *cul-de-sac* in a problem-solving sequence and recognizes he has not solved the problem is likely to become more anxious than he was initially. As a result of the increased anxiety his selection and evaluation of a second solution path is apt to be impaired, and the probability of success attenuated. This maladaptive cycle may become entrenched with time and after 5 years of experiencing the sequence: problem ⟶ impulsive selection of invalid solution sequence ⟶ failure ⟶ anxiety ⟶ selection of second sequence . . . ⟶ failure . . . etc., the child may gradually withdraw involvement from problem situations and become apathetic or hostile toward intellectual situations.

Relation to Basic Research and Theory

The points of contact between reflection and analysis and the main stream of psychological theory in cognition are still unclear. Witkin and his colleagues (1962) have used solution time on the EFT as an index of the construct of field independence. The data summarized indicate no strong relation between field independence and reflection impulsivity. Experimental psychologists have noted recently that discrimination response time is not only a function of information transmitted (i.e., stimulus uncertainty), but is also determined by response uncertainty (Morin & Forrin, 1963; Seibel, 1963). It is believed that further investigations of response uncertainty will provide an important bridge between the traditionally isolated domains of cognition and personality. For available data suggest that individual-difference variables like motivation, expectation of failure, and hyperkinesis are responsible, in part, for the dramatic variability associated with laboratory studies of response time, perceptual dynamics, and human learning. The data generated by studies in these three substantive areas are influenced, in varying degrees, by *S*'s characteristic strategy in situations with high-response uncertainty.

Theoretical conceptualizations of intellective development have been mainly concerned with the growth and dynamics of mediational systems and cognitive operations (e.g., associational transformations, algorithms, rules of reasoning). Piaget's theorizing, for example, has centered on the characteristics of cognitive structures at different stages of development, and Piaget assumes that differences in these structures explain age differences in cognitive productions. Both European and American psychologists have typically been indifferent to the cognitive phenomena that describe the beginning and end of a problem-solving sequence: the initial coding of information, the selection of a "best" solution hypothesis, and the psychological characteristics of the response the child uses to communicate his answer to a social agent. The reflection and analysis variables are essential aspects of these processes.

This paragraph is an attempt to dilute the authors' tendency to imply that the reflective child was necessarily the better or brighter child. It seems reasonable to assume that efficient learning and performance on varied intellective tasks will sometimes be facilitated by a reflective or analytic approach; sometimes by a more impulsive or less analytic orientation. Indeed, a recent study (Lee et al., 1963) demonstrated this last point clearly. Some of the academic contents children must master require reflection and analysis especially the disciplines with deductive structures such as mathematics and physical science. But mastery and productivity in the humanities, arts, and social sciences can be hampered by an excessively strong reflective orientation. New pedagogical procedures should acknowledge the interactions between the dispositions of the learner and the material, and tailor presentations to the preferred strategy of the child.

SUMMARY

These studies have uncovered two basic cognitive dispositions that appear to have generality and stability: the tendency to analyze visual arrays and the tendency to reflect upon alternative-solution hypotheses in situations where the alternative-response possibilities are available simultaneously. Each of these dispositions contributes considerable variance to the production of the more complex cognitive products of analytic concepts on CST and recognition errors on tasks requiring complex perceptual discriminations (DRT, MFF). The studies upon which these conclusions are based are summarized below.

Study 1

Investigation of first and second grade boys revealed that the production of analytic concepts on CST was positively related to the tendency to inhibit incorrect solution hypotheses on EFT, high scores on the Picture Arrangement subtest of WISC, and the production of articulated ink-blot percepts. However, analytic concepts were unrelated to verbal skills. These findings linked analytic concepts with the tendency to reflect over alternative response possibilities.

Study 2

Second grade children were assigned to one of two experimental groups and administered CST, DRT, and EFT. One group was instructed to respond quickly; the other group was instructed to respond slowly and to reflect upon their answers. Children in the slow-instruction group produced more analytic concepts on CST and made fewer errors on DRT, suggesting that experimental manipulation of the tendency to reflect upon alternative solutions does affect analytic concepts and perceptual-recognition errors in the manner predicted.

Study 3

Second and third grade children were administered CST, DRT, and several related tests in two sessions separated by 9 weeks. There was remarkably high consistency of recognition error and response-time scores across the two administrations on DRT that were separated by 9 weeks. Moreover, there was an inverse relationship between recognition errors and response time. Finally, analytic concepts on CST were inversely related to error scores on DRT and related perceptual discrimination tests.

Study 4

A group of 57 fourth grade boys was administered CST and a measure of visual analysis. There was a positive correlation between analytic concepts and the tendency to analyze the geometric stimuli into distinct figural and background components, indicating that visual analysis was a determinant of analytic groupings.

Study 5

Third and fourth grade *S*s who were seen in Study 3 were administered new tests of visual analysis and reflection impulsivity 1 year later. These scores were correlated with the earlier measures of reflection impulsivity. There was moderately high stability for the reflection-impulsivity dimension over the 1-year period. For response time and error scores on DRT task were related to response time and errors on MFF 1 year later (correlations in the .40's). Correct recognition of the element components in the visual-analysis task was positively related to analytic concepts on CST, negatively related to error scores on DRT or MFF, but independent of response time or verbal ability. When visual-analysis scores and response time were used as individual predictors, and analytic concepts on CST or error scores on MFF or DRT were the criterion, the multiple correlations were much higher than those obtained when only a single predictor was used.

Study 6

This study assessed the effect of impersonal versus reassuring testing conditions on the reflection-impulsivity dimension among third grade children. There was no consistent difference in errors or response time on three different tasks between *Ss* examined under "cold versus warm" testing conditions. However, there was marked intraindividual consistency in response time and errors across the three tasks, supporting the generality of this cognitive disposition.

Study 7

A small group of Fels children (7–11 years of age) was administered a battery of tests including CST. The results supported the findings from Studies 1, 2, and 3, for analytic concepts were inversely correlated with the occurrence of incorrect hypotheses on perceptual discrimination tests.

Study 8

Several independent analyses assessed the relationship between analytic concepts on CST and the behavioral variables of task unrelated gross motor behavior and distractibility. In one study analytic concepts were positively correlated with the ratio of task attention to distractibility in the classroom. Moreover, analytic boys were less likely than nonanalytic boys to display bursts of gross motor behavior on a school playground. Further, production of analytic concepts was positively correlated with regularity of respiration under conditions of rest and during attention to simple stimuli. Finally, boys who were analytic, in contrast to nonanalytic *Ss*, during the early school years were rated as less hyperkinetic and more involved in solitary task mastery during the first 8 years of life.

The accumulated evidence suggests that reflection over alternative-solution possibilities and visual analysis are fundamental cognitive dispositions that influence both analytic concepts on CST and perceptual recognition errors on DRT or MFF. It was suggested that the determinants of these two dispositions may be related to constitutional variables, degree of involvement in the task, and anxiety over task competence.

REFERENCES

FISHER, S., & CLEVELAND, S. E. *Body image and personality.* Princeton: Van Nostrand, 1958.

GETZELS, J. W., & JACKSON, P. W. *Creativity and intelligence.* New York: Wiley, 1962.

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist,* 1959, **14**, 469–479.

INHELDER, BÄRBEL, & PIAGET, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

KAGAN, J. Developmental studies in reflection and analysis. In A. H. Kidd & J. H. Rivoire (Eds.), *Perceptual and conceptual development in children.* N. Y.: International Universities Press, in press.

KAGAN, J., & MOSS, H. A. The availability of conflictful ideas: A neglected parameter in projective test responses. *J. Pers.,* 1961, **29**, 217–234.

KAGAN, J., & MOSS, H. A. *Birth to maturity.* New York: Wiley, 1962.

KAGAN, J., MOSS, H. A., & SIGEL, I. E. Psychological significance of styles of conceptualization. *Monogr. Soc. Res. Child Develpm.,* 1963, **28**(Ser. No. 86, Whole No. 2) 73–112.

KAGAN, J., & ROSMAN, BERNICE L. Autonomic correlates of attention and an analytic attitude. *J. exp. child Psychol.,* in press.

LEE, L. C., KAGAN, J., & RABSON, ALICE. The influence of a preference for analytic categorization upon concept acquisition. *Child Develpm.,* 1963, **34**, 433–442.

MORIN, R. E., & FORRIN, B. Response equivocation and reaction time. *J. exp. Psychol.,* 1963, **66**, 30–36.

ROSMAN, BERNICE L. Analytic cognitive style in children. Unpublished doctoral dissertation, Yale University, 1962.

SCHAEFER, E. S., & BAYLEY, NANCY. Maternal behavior, child behavior and their intercorrelations from infancy through adolescence. *Monogr. Soc. Res. Child Develpm.,* 1963, **28**(Ser. No. 87, Whole No. 3).

SEIBEL, R. Discrimination reaction time for a 1023 alternative task. *J. exp. Psychol.,* 1963, **66**, 215–226.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation.* New York: Wiley, 1962.

WITKIN, H. A., LEWIS, HELEN, B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL, & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

WOHLWILL, J. F. Developmental studies of perception. *Psychol. Bull.,* 1960, **57**, 249–288.

(Received June 14, 1963)

Vol. 78, No. 2 Whole No. 579, 1964

Psychological Monographs: General and Applied

IDIODYNAMIC SETS IN WORD ASSOCIATION[1]

LOUIS J. MORAN
University of Texas

ROY B. MEFFERD, Jr.
Central Veterans Administration Psychiatric and Psychosomatic Research Laboratory, Houston, Texas

AND JAMES P. KIMBLE, Jr.
University of Texas

On 4 successive days 100 normals and 96 schizophrenics were administered a list of 125 words under free-association instructions. From these samples 79 normal-schizophrenic pairs were matched for age, education, and vocabulary. Separate orthogonal factor anlyses of each day's performance by 79 normal Ss yielded essentially the same factors. 3 factors were interpreted as representing idiodynamic sets: object-referent, conceptual-referent, and speed. The manner in which these sets facilitated and interfered with mediational processes was demonstrated. Similar analyses of performance by the 79 schizophrenics revealed the same 3 idiodynamic sets. Although the sets were operative in the schizophrenics on all 4 days, the sets were less pre-emptive after the 1st day.

In the controlled word-association experiment subjects are provided a "set" as part of the task, e.g., "Give a synonym of ———." No set is provided by the examiner in the free word-association experiment, e.g., "Tell me the first word you think of ———." Even in free word association, however, many subjects respond as though they were under controlled association instructions. Such subjects consistently give one type of response, e.g., contrast, wherever possible. Furthermore, this "preference" for a specific kind of associative response seems to be an enduring characteristic of the individual. For example, Karwoski and Berthold (1945) report a test-retest correlation of .92 for contrast responses with a 40-day interval between tests.

Numerous investigators have commented on what might be called idiodynamic sets in word association. Of particular interest has been the contrast versus the defining set (Karwoski & Berthold, 1945; Kent & Rosanoff, 1910; O'Connor, 1934; Tendler, 1933; Wells, 1919). Scientists reportedly differ from "literary" types in the tendency to give contrast responses (Murphy, 1917). Men reportedly give more contrasts, and women give more defining responses (Goodenough, 1946). According to Karwoski and Berthold (1945),

> contrast categories and the similarity categories are distinctive and probably represent different mental processes. Combined with the fact that the reliability figures from grouped similarity and grouped contrast are very high, the indication is that there is a tendency on the part of individuals to respond either in contrasts or similarities [p. 99].

To the extent that an individual is consistent in the use of one or another specific principle (e.g., contrast or similarity) in mediating stimulus-response (SR) word connections, one may speak of his characteristic associative set. The present study seeks to corroborate the findings of other investigators regarding individual differences in characteristic associative set; to describe more fully a few specific sets; and, most importantly, to examine the implications of idiodynamic set for the study of association structure by means of word association.

---

[1] This study is part of a continuing project of the Psychiatric and Psychosomatic Research Laboratory, Houston Veterans Administration Hospital, and was conducted in Houston, Texas, under the auspices of that laboratory. A version of the present paper was presented at the Third Annual Meeting of the Society for Multivariate Experimental Psychology, Chicago, 1962. Financial assistance was given by the Hogg Foundation for Mental Health and the University of Texas Research Institute.

METHOD

Subjects

Initially, 100 normal men and 96 schizophrenic men were examined. The normal subjects were drawn from many sources: nonpsychiatric patients, hospital aides, janitors, firemen, etc. The schizophrenic subjects were mostly acutely psychotic, tested as soon after admission as circumstances permitted; some subjects were receiving chemotherapy; others not. Normal and schizophrenic subjects then were matched by pair for age, education, and estimated intelligence (Wechsler-Bellevue—WB—Vocabulary subtest, Form I). This procedure yielded 79 closely matched pairs. The matched-normal sample averaged .6 years younger, were of identical education level, and scored an average of .2 raw score points lower on WB Vocabulary than the schizophrenic subjects (cf. Tables 1 and 6, Variables 13–15).

Word Lists

The total list of 500 familiar words consisted of 20 separate forms, 25 words each, roughly equated for Commonality (sum frequency of three most common responses), grammatical form, and homonym type. Each form contained 5 words from the Kent-Rosanoff list, 13 homonyms, 2 neutral words, and 5 Kent-Rosanoff words repeated from the preceding form. The complete word lists and a detailed account of their construction have been published elsewhere (Moran, Mefferd, & Kimble, 1960a).

Test Administration

The subjects were instructed to

> tell me the first word that comes to mind when you hear this word.

After each form (25 words), the examiner stopped and obtained reproductions, i.e.,

> Now I am going to say these same words again, and I would like for you to give me the same word you gave the first time.

The subjects were urged to respond rapidly and were timed with a stop watch.

The subjects were administered five forms (125 words) on each of 4 successive days. Forms were so arranged that on the last 3 days, the subject responded to 25 words he had received on the preceding day (the measure thus derived is not included in this analysis due to a very high positive relationship to immediate reproduction). All subjects were tested individually. The Vocabulary test was administered at the end of the fourth day's testing. Testing was accomplished by 15 different examiners.

Variables

The 15 major variables used in the present study are listed in Table 1. Variables 1–6 may be termed structural variables selected to characterize individual differences in association structure. They are very similar to those employed by Karwoski and Berthold (1945) and many others interested in idiodynamic associative sets.[2] Consistent patterns of relationship among these structural variables provide the primary evidence for set in the present study. Variables 7–11 represent mediation faults from which is inferred the current mediational efficiency of association structure. These variables are traditional "complex indicators." The last four variables (12–15) are descriptive of more general characteristics of the subjects. Later in the analysis, 9 minor variables were introduced; these variables (16–24) are described briefly in Table 4.

1. Synonym. The response word (RW) has exactly the same meaning as the stimulus word (SW) in one or more ordinary and appropriate contexts, e.g., BLOSSOM, FLOWER.

2. Contrast. RW negates or contrasts with the meaning of SW in one or more ordinary and appropriate contexts, e.g., DARK, LIGHT.

3. Logical Coordinate. SW and RW separately denote immediate members (of equal logical order) of the same class or category, e.g., BLUE, YELLOW. Logical Coordinates that meet also the criteria for Contrast are scored Contrast.

4. Superordinate. SW denotes an immediate member of the class or category denoted by RW, e.g., CABBAGE, VEGETABLE.

5. Subordinate. RW denotes an immediate member of the class or category denoted by SW, e.g., VEGETABLE, CABBAGE.

6. Functional. SW and RW each separately denote entities or processes between which there is an explicit functional relationship, e.g., NEEDLE, THREAD. This category also has priority over Logical Coordinate, e.g., KNIFE, FORK, and is scored Functional.

7. Multiword. The response consists of more than one word, e.g., MALE, A MAN.

8. Blank. Subject fails to respond within 20 seconds or indicates that he has no response. Also included in this category are repetitions of SW: RW is exactly the same word as SW.

9. Distant. RW has no relation or a very remote relationship to SW. A lengthy narrative would be required to relate SW and RW, e.g., FAITH, BOX.

10. Reaction Time. Average reaction time, in seconds.

11. Reproduction Failure. The number of times

---

[2] Detailed instructions for scoring the structural variables and tables giving the frequency of each kind of response (Variables 1–12) to each of the 400 different stimulus words and unrotated factor loadings have been deposited with the American Documentation Institute. Order Document No. 7782 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $2.00 for microfilm or $3.75 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

the subject failed to give exactly the same associates, when retested with the same word list.

12. Commonality. Each of the subject's initial responses (not including Reproduction) was assigned a value corresponding to the percentage of times that response was given by the total sample of 196 subjects. Commonality scored in this manner correlated .90 with Commonality scored according to the Minnesota norms (Russell & Jenkins, 1954) for the 100 Kent-Rosanoff words.

13. Age. Number of years.

14. Education. Number of years formal education.

15. Vocabulary. WB (Form I) Vocabulary raw score.

Statistical Analysis

Pearson product-moment correlation coefficients, using raw scores,[3] were factored into principal components and rotated by the normalized varimax method. Unities were placed in the diagonal. Factor extraction was stopped when eigen values dropped below unity.

RESULTS ON NORMAL SUBJECTS

The means, standard deviations, and intercorrelations of the 15 variables on each day and total days are provided in Table 1. The rotated factor loadings on each of the 4 successive days, shown in Table 2, were sufficiently similar to be discussed collectively, at least for the present.

Factor V, when it emerged, was characterized mainly by the Subordinate and/or Multiword variables. The h^2 values of .39 for Subordinate and .36 for Multiword (Table 2, total 4 days) indicate that these two variables were minimally related to the other variables in the analysis. Factor V is a residual factor of little interest. On Factor I, which might be labeled "general ability to cope with the task," Education and Vocabulary consistently loaded high positive, with all of the mediational faults, and Age, loading negatively. Factors V and I do not bear directly upon idiodynamic set.

The clearest evidence for the operation of idiodynamic sets appeared in Factors II, III, and IV on each day. These factors, characterized chiefly by the structural variables, indicate that Functional, Synonym-Superordinate, and Contrast-Coordinate were three essentially independent modes of responses. When results of the 4 days were pooled, and all 500 associations were included, the same factors appeared (Table 2, total 4 days).

The great similarity of each set factor to its corresponding set factor on each of the 4 successive days may be determined by inspection of the loadings in Table 2. To obtain more objective factor matching, however, the following procedure was adopted. The factor scores of each subject on each factor on all days were calculated yielding 19 factor scores for each subject. These factor scores were treated as variables (i.e., the five factors on Day 1 as Variables 1–5; the four factors on Day 2 as Variables 6–9, etc.) and factor analyzed. The results are given in Table 3.

As may be seen in Table 3, Factors I–IV were essentially the same on all days, with the residual Factor V emerging as two independent residuals on different days. Since the factor scores of the subjects were used in this analysis, the results given in Table 3 establish also the high reliability of subjects with respect to idiodynamic set over the 4 days. Another important point is that only 3 of the 79 subjects had a factor score greater than 1.00 on more than one of the set factors (II–IV). This datum indicates that the subjects may be sorted into distinctive associative types, as suggested by Karwoski and Berthold (1945) and others.

It remains now to attempt a more complete description of these three idiodynamic sets and to explore their implications for the understanding of association structure. For the purpose of describing the sets three sources of information were available in the present study: the distribution of loadings on the factors, the relationship of other response tendencies to the major variables of each set, and qualitative observations from the protocols of subjects with different sets. Only the broad outlines of each set factor will be mentioned in this section, with details and interpretations reserved for the Discussion section.

From inspection of the distribution of loadings on Factor II, it would appear that the

---

[3] The entire analysis was repeated using square root transformed scores to correct for the skewness of some variables. Differences between the two solutions were negligible. All computations were carried out at the Computation Center of Southern Methodist University, with the assistance of P. Minton, B. McCauley, and J. B. Harvill.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF 15 VARIABLES FOR 79 NORMAL MEN ON 4 SUCCESSIVE DAYS

Variable	*M*	*SD*	Day 1 intercorrelations													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Synonym	16.1	7.4														
2. Contrast	10.3	6.7	08													
3. Coordinate	8.7	5.1	12	62												
4. Superordinate	12.3	5.6	42	−18	−04											
5. Subordinate	4.9	2.4	34	−06	−05	−08										
6. Functional	19.0	6.2	−15	06	−38	−05	−09									
7. Multiword	5.1	7.2	−10	−24	−14	08	02	−26								
8. Blank	3.4	5.0	−34	−12	06	−14	−09	−20	14							
9. Distant	4.3	3.9	−51	−26	−06	−37	−23	−36	−08	30						
10. Reaction Time (seconds)	2.9	1.2	−37	−25	−20	−01	−21	−13	27	58	35					
11. Reproduction Failure	19.7	20.1	−32	−30	−10	−19	−18	−34	19	37	53	38				
12. Commonality (%)	12.1	2.9	49	64	34	16	19	35	−26	−37	−60	−47	−52			
13. Age	34.7	8.3	−13	08	−15	14	−01	21	17	24	−11	29	09	01		
14. Education	10.9	3.4	42	19	05	−02	20	18	−31	−53	−37	−40	−23	40	−25	
15. Vocabulary	20.1	6.7	56	33	10	13	23	17	−27	−48	−54	−44	−43	48	−02	66
			Day 2 intercorrelations													
1. Synonym	11.2	5.5														
2. Contrast	9.7	6.4	04													
3. Coordinate	9.1	4.9	17	51												
4. Superordinate	12.5	6.8	55	−02	01											
5. Subordinate	3.2	1.9	32	13	19	03										
6. Functional	21.1	6.7	−19	02	−27	−22	−21									
7. Multiword	4.2	6.9	−07	−32	−14	04	−19	−19								
8. Blank	1.9	3.1	−23	−13	−06	−13	−18	−16	14							
9. Distant	3.8	4.1	−46	−18	−07	−33	−28	−29	05	25						
10. Reaction Time (seconds)	2.7	1.3	−14	−33	−26	−04	−22	−10	25	34	32					
11. Reproduction Failure	18.2	17.2	−39	−31	−04	−18	−28	−34	29	26	49	31				
12. Commonality (%)	12.8	3.1	33	63	28	19	24	32	−44	−24	−57	−47	−61			
13. Age	34.7	8.3	16	08	06	29	−08	15	09	20	−08	26	−06	08		
14. Education	10.9	3.4	21	12	08	−12	31	08	−33	−39	−28	−36	−26	26	−25	
15. Vocabulary	20.1	6.7	47	25	06	17	36	17	−23	−44	−52	−38	−47	37	−02	66
			Day 3 intercorrelations													
1. Synonym	10.8	6.3														
2. Contrast	9.7	6.7	11													
3. Coordinate	11.0	3.7	19	48												
4. Superordinate	12.2	6.0	45	−05	−10											
5. Subordinate	2.8	1.7	27	19	14	06										
6. Functional	18.7	5.5	−00	09	−08	−03	−15									
7. Multiword	4.4	7.8	−12	−22	−19	14	−15	−21								
8. Blank	1.6	3.1	−19	−13	−13	−21	−00	−04	07							
9. Distant	4.8	5.2	−45	−15	−10	−44	06	−51	−08	25						
10. Reaction Time (seconds)	2.6	1.2	−13	−40	−35	01	−08	−12	11	33	18					
11. Reproduction Failure	20.4	16.8	−28	−20	−12	−20	13	−39	19	31	63	18				
12. Commonality (%)	12.4	2.8	39	62	26	32	07	50	−34	−36	−61	−42	−54			
13. Age	34.7	8.3	09	09	03	16	−23	13	05	13	−14	20	06	11		
14. Education	10.9	3.4	31	05	26	−04	08	10	−16	−42	−24	−32	−30	23	−25	
15. Vocabulary	20.1	6.7	56	18	30	22	15	12	−21	−48	−44	−34	−40	42	−02	66
			Day 4 intercorrelations													
1. Synonym	13.5	6.3														
2. Contrast	7.8	5.2	16													
3. Coordinate	10.3	4.9	29	44												
4. Superordinate	13.2	6.9	54	04	−00											
5. Subordinate	4.9	2.4	28	25	26	−01										
6. Functional	21.4	6.3	−16	02	−35	04	−09									
7. Multiword	3.0	4.9	00	03	−03	13	−07	−02								
8. Blank	1.4	2.9	−22	05	03	−15	−06	−01	01							
9. Distant	3.6	3.9	−38	−20	−04	−19	−26	−39	−12	21						
10. Reaction Time (seconds)	2.6	1.2	−02	−29	−10	09	−25	−12	04	22	28					
11. Reproduction Failure	15.2	14.5	−29	−13	08	−16	−24	−37	−02	25	60	28				
12. Commonality (%)	13.2	2.8	36	56	15	37	20	48	−10	−23	−60	−37	−47			
13. Age	34.7	8.3	12	15	05	18	14	14	12	14	−08	20	−05	18		
14. Education	10.9	3.4	18	07	00	−17	11	04	−19	−40	−25	−38	−26	18	−25	
15. Vocabulary	20.1	6.7	39	19	04	04	22	04	−13	−42	−40	−34	−38	25	−02	66

Table 1—Continued

Variable	*M*	*SD*	Total intercorrelations (4 days)													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Synonym	51.6	22.6														
2. Contrast	37.6	23.0	14													
3. Coordinate	39.0	15.0	24	62												
4. Superordinate	50.1	22.8	56	—04	—00											
5. Subordinate	15.8	6.3	42	14	23	01										
6. Functional	80.2	21.4	—16	05	—30	—08	—17									
7. Multiword	16.7	23.8	—08	—25	—13	13	—13	—18								
8. Blank	8.3	12.5	—26	—12	—04	—18	—13	—13	15							
9. Distant	16.4	15.8	—51	—22	—06	—36	—26	—47	—03	32						
10. Reaction Time (seconds)	2.7	1.2	—20	—34	—27	00	—27	—13	19	44	30					
11. Reproduction Failure	87.9	60.8	—42	—23	—05	—16	—22	—39	19	35	63	34				
12. Commonality (%)	12.6	2.7	41	66	31	26	21	47	—32	—33	—66	—45	—58			
13. Age	34.7	8.3	06	11	—01	22	—04	18	12	21	—11	25	12	10		
14. Education	10.9	3.4	32	12	11	—10	23	11	—28	—50	—30	—38	—37	29	—25	
15. Vocabulary	20.1	6.7	56	26	14	15	32	14	—24	—51	—51	—39	—53	42	—02	66

TABLE 2

ROTATED ORTHOGONAL FACTOR LOADINGS ON 4 SUCCESSIVE DAYS

Variable	I	II	III	IV	V	h^2
Day 1						
1. Synonym	44	−12	62	14	39	77
2. Contrast	13	19	−09	91	−02	88
3. Coordinate	03	−38	03	85	−06	88
4. Superordinate	−02	01	90	−10	−18	85
5. Subordinate	13	−03	−04	−04	90	84
6. Functional	21	90	−12	−12	−15	90
7. Multiword	−51	−15	30	−21	24	48
8. Blank	−76	−12	−21	07	00	64
9. Distant	−26	−47	−57	−23	−32	77
10. Reaction Time (seconds)	−72	−02	−08	−21	−12	59
11. Reproduction Failure	−41	−41	−30	−27	−08	51
12. Commonality (%)	41	40	27	62	16	80
13. Age	−58	53	14	08	15	67
14. Education	75	10	03	06	24	63
15. Vocabulary	61	23	26	23	32	65
% variance (72)	22	13	13	15	9	
Day 2						
1. Synonym	28	−12	82	06		76
2. Contrast	02	04	03	88		77
3. Coordinate	00	−44	07	72		72
4. Superordinate	−09	−11	81	−05		68
5. Subordinate	50	−26	27	20		43
6. Functional	01	92	−13	04		87
7. Multiword	−29	−22	13	−50		40
8. Blank	−62	−14	−14	−04		42
9. Distant	−33	−38	−60	−17		64
10. Reaction Time (seconds)	−53	−08	−01	−45		49
11. Reproduction Failure	−30	−49	−41	−34		62
12. Commonality (%)	21	41	33	71		83
13. Age	−56	21	47	10		59
14. Education	81	09	−01	10		67
15. Vocabulary	69	22	40	13		71
% variance (63)	19	12	16	16		

Table 2—Continued

Variable	I	II	III	IV	V	h^2
Day 3						
1. Synonym	24	08	76	10	32	75
2. Contrast	−11	17	02	85	17	80
3. Coordinate	16	−10	03	76	10	62
4. Superordinate	−03	06	83	−08	−11	72
5. Subordinate	06	−29	23	11	74	70
6. Functional	−04	87	−09	−02	01	76
7. Multiword	−06	−37	21	−15	−69	68
8. Blank	−64	−12	−20	−21	18	54
9. Distant	−18	−64	−56	−09	16	79
10. Reaction Time (seconds)	−43	−09	06	−62	06	59
11. Reproduction Failure	−33	−69	−22	−09	03	64
12. Commonality (%)	10	65	33	52	15	85
13. Age	−56	16	32	17	−26	54
14. Education	82	13	05	10	11	71
15. Vocabulary	66	20	43	23	16	73
% variance (69)	15	16	15	14	9	
Day 4						
1. Synonym	30	06	79	31	02	82
2. Contrast	−09	27	−06	78	−05	70
3. Coordinate	−02	−27	12	78	−02	70
4. Superordinate	−07	18	81	−04	−13	71
5. Subordinate	14	11	10	56	11	37
6. Functional	−14	84	−16	−27	08	82
7. Multiword	−12	02	07	−03	−95	92
8. Blank	−67	−10	−23	10	09	54
9. Distant	−29	−70	−23	−18	17	68
10. Reaction Time (seconds)	−47	−33	36	−38	12	62
11. Reproduction Failure	−33	−67	−17	−06	07	60
12. Commonality (%)	07	76	23	40	07	80
13. Age	−48	26	35	15	09	45
14. Education	81	12	−13	09	18	73
15. Vocabulary	74	22	17	19	15	68
% variance (67)	16	18	12	14	7	
Total (4 days)						
1. Synonym	34	−02	82	20		82
2. Contrast	03	22	−02	89		84
3. Coordinate	05	−25	10	84		79
4. Superordinate	−12	10	81	−07		68
5. Subordinate	37	−21	35	30		39
6. Functional	04	92	−18	−12		89
7. Multiword	−35	−19	30	−34		36
8. Blank	−68	−17	−17	01		52
9. Distant	−28	−62	−55	−11		78
10. Reaction Time (seconds)	−58	−15	01	−38		51
11. Reproduction Failure	−48	−54	−29	−12		63
12. Commonality (%)	24	65	26	57		88
13. Age	−58	30	31	12		54
14. Education	81	10	04	06		67
15. Vocabulary	68	23	38	16		69
% variance (66)	20	16	15	15		

TABLE 3

ROTATED ORTHOGONAL FACTOR LOADINGS OF FACTOR SCORES

Variable	Day	Factor	I	II	III	IV	V	VI	h^2
1	1	I	94	09	11	10	10	09	94
2	1	II	08	87	10	03	05	14	70
3	1	III	12	03	84	01	27	18	81
4	1	IV	04	03	08	82	08	08	83
5	1	V	13	09	14	03	02	95	95
6	2	I	97	01	02	02	07	12	96
7	2	II	01	88	15	17	01	09	91
8	2	III	02	22	88	07	08	28	91
9	2	IV	01	19	08	92	15	02	83
10	3	I	97	04	05	06	08	01	82
11	3	II	07	88	01	06	19	04	82
12	3	III	06	06	93	06	01	16	95
13	3	IV	07	03	07	89	15	03	90
14	3	V	11	22	05	15	84	15	82
15	4	I	97	02	11	03	00	08	86
16	4	II	05	87	18	20	12	13	96
17	4	III	13	08	90	11	10	07	86
18	4	IV	00	16	09	87	05	27	87
19	4	V	02	15	11	02	87	11	81

Note.—Signs omitted because of arbitrary sign reversals occurring in factor scores derived from four separate factor analyses.

only feature sharply distinguishing this set factor from the other two set factors was the dominance of the Functional variable and independence of the other structural variables (except for a slight negative relationship to Coordinate). The positive loading of Commonality and the negative loading of Reproduction Failure in Factor II also appeared, in differing magnitudes, on the other two set factors. The negative loading of Distant associates in Factor II also appeared on Factor III.

The relationship of Functional associates to other response tendencies, shown in Table 4, was especially helpful in differentiating Factor II from the other set factors. Functional associates typically involved a relationship between concrete objects, e.g., FOOT, SHOE; DOCK, BOAT. It may be noted in Table 4 that the subjects who gave many Functional associates tended also to name a concrete object in response to SWs that denoted an attribute (Variables 18–19) or an action (Variable 20). The object-referent associates typical of the subjects who gave many Functional associates resulted in many noun responses (Variable 22). It may be seen in Table 4 (Variables 18–20 and 22) that all of these tendencies, characteristic of the subjects with a set to give Functional associates, were reversed (negatively correlated) with all of the other structural variables.

The two structural variables characterizing Factor III, essentially independent of the other structural variables, were Synonym and Superordinate. As with the other two set factors, Commonality loaded positively, and Reproduction Failure loaded negatively. Factor III shared with Factor II a negative loading on Distant associates. Age and Education seemed more strongly related to Factor III than to the other set factors.

TABLE 4

CORRELATES OF THE STRUCTURAL VARIABLES

Other response variable	Structural variable				
	Functional	Synonym	Superordinate	Contrast	Coordinate
16. RW intrinsic predicate of SW (e.g., APPLE, RED)	00	−10	04	−29	−37
17. RW extrinsic predicate of SW (e.g., APPLE, ROTTEN)	−05	−37	−44	−21	−30
18. SW intrinsic predicate of RW (e.g., RED, APPLE)	30	−20	−17	−60	−52
19. SW extrinsic predicate of RW (e.g., ROTTEN, APPLE)	37	−37	−27	−56	−58
20. Action of or upon RW (e.g., HIT, BALL)	45	−50	−44	−35	−49
21. Associated actions (e.g., RUN, LEAP)	−12	59	39	40	36
22. Total noun responses	40	−37	−13	−33	−31
23. Total verb responses	−19	62	60	15	11
24. Total adjective responses	−23	20	−11	52	42

Note.—Variables 16–21 and the 5 structural variables represent mutually exclusive categories, i.e., no response received more than 1 of these 11 scores. Grammatical form of RW, Variables 22–24, was not scored for ambiguous forms, e.g., noun-verbs, noun-adjectives, etc. Pearson *r's* were computed for total performance (on 500 SWs) of the 79 subjects.

From Table 4 it may be seen that, in sharp contrast with Functional associates, Synonym-Superordinate associates were negatively correlated with object-referent responses to adjective and verb SWs (Variables 18–20) and with noun responses, per se (Variable 22). The "concretizing" type of associate which names a possible attribute of the object denoted by SW also was negated by the set to give Synonym-Superordinate associates (Variable 17). Instead of naming a concrete object, subject to the action denoted by a verb SW (Variable 20), the subjects giving Synonym-Superordinate associates tended to give verb responses to verb SWs (Variable 21) and to give many verb responses in general (Variable 23).

Contrast-Coordinate were the dominant variables on Factor IV, independent of the other structural variables. Commonality also loaded positively. Average Reaction Time tended to load negatively, indicating that the subjects high on Factor IV tended to give rapid responses. The pattern of correlations with other response tendencies was very similar for the Synonym-Superordinate and the Contrast-Coordinate variables, with one major exception, as shown in Table 4: the subjects giving Contrast-Coordinate associates had a strong tendency to give adjective responses (Variable 24) rather than verb or noun responses.

Table 5 was designed to illustrate the importance of taking into account idiodynamic sets when interpreting responses in the word-association experiment.

As shown in Table 5, the greatest number of high Commonality responses was given to SWs that "fit" the subject's set. Conversely, the lowest number of "individual" responses (i.e., given by less than 3% of the population) was given to SWs most appropriate to the subject's set. Frequency of mediation faults (Variables 7–11, with Reaction Time greater than 3 seconds counted as a fault) also was predictable (except for Factor III subjects on Functional SWs) from knowledge of the subject's set.

Discussion

In the word-association experiment, the term *set* traditionally has been reserved to denote a special orientation of the subject, induced by the examiner's instructions or experimental conditions. All subjects in the present experiment received the same instructions from the examiner, yet three groups of

TABLE 5

Effect of Set Upon Commonality and Mediation Faults

Type of set	Type of SW		
	Functional	Synonym-Superordinate	Contrast-Coordinate
Average commonality (%) of responses			
II. Functional	19.3	14.5	16.0
III. Synonym-Superordinate	13.6	15.3	10.3
IV. Contrast-Coordinate	16.1	13.5	19.8
Average frequency of individual responses			
II. Functional	3.8	6.0	6.0
III. Synonym-Superordinate	7.2	5.6	9.4
IV. Contrast-Coordinate	7.2	7.7	5.3
Average frequency of mediation faults			
II. Functional	5.7	9.6	8.3
III. Synonym-Superordinate	11.0	11.0	12.5
IV. Contrast-Coordinate	9.0	11.0	7.2

Note.—Each type of set is represented by 12 subjects with factor score greater than .50 on one set factor and less than .50 on the other set factors ($N = 36$). Each type of SW is represented by 20 SWs that elicited predominately one kind of response (Functional, Synonym-Superordinate, or Contrast-Coordinate) to the exclusion of the other two kinds ($N = 60$ words). The three groups of SWs were equated (from norms on 196 subjects for sum frequency of primary responses, i.e., Commonality value. Commonality of responses is the percentage of times each subject's response was given by 196 subjects. Individual response is given by less than 3% of 196 subjects. Mediation faults are all faults (can exceed one per word, e.g., a delayed Reaction Time, Distant response, with Reproduction Failure = three faults).

subjects behaved as if under quite different, but very specific, instructions. Since these subjects, on different occasions and with different word lists, were highly consistent in following their own implicit instructions, or personal interpretation of the examiner's instructions, they may be said to have evidenced three different sets. Although no inquiry was conducted to confirm the point, it is highly unlikely that these subjects were aware of their set. Investigators from the Wurzberg school to the present time have noted the inability of the subjects to describe the processes mediating SR word connections in the word-association experiment.

Adult subjects in controlled association experiments experience little difficulty in adopting a set to give either Functional, or Synonym, or Contrast responses. What, then, determines the tendency of many subjects to adopt predominately one or the other of several equally "possible" sets in the free-association experiment? In the present study differences in "attitude toward words" appeared to parallel differences in the subjects' implicit instruction. It may be hypothesized that the subject's basic attitude toward words influenced his interpretation of the examiner's instruction. When asked to associate words an individual with a concrete, denotative attitude toward words would tend to give object-referent associates. An individual with a more abstract, conceptual attitude toward words would be more likely to associate words by their logical relationships to other words. These differences in attitude toward words would give rise to idiodynamic associative sets resembling Factors II and III, for example.

It seems highly probable that in Factor II the set was to interpret words as denoting concrete objects. Associations tended to be object-object, attribute-object, action (on)-object, etc. The SR word connections were based upon experiential contiguity, not "logical" relationships. The high-positive loading of Commonality suggests that conventionality of object-referent associates, connected by experiential contiguity, was a major source of variability on this factor.

The negative loading of Distant associates on both Factors II and III seemed inconsistent with an interpretation of Factor II as representing an object-referent set. Inspection of the Distant associates given by Factor III subjects revealed them to be conceptually remote from SW, as the name of the category implies, e.g., PROBE, POSSIBLE; PROMISE, LOSS. The so-called Distant associates of Factor II subjects, however, commonly were really concrete, trivial aspects of the thing denoted by SW, e.g., SCAB, HAIR. Although both types of responses were assigned to the same category in the present study, the reanalysis revealed a qualitative difference between Distant associates on the two factors. The Distant associates of Factor II subjects were consistent with an object-referent set.

In contrast to Factor II, the Factor III subjects seemed to interpret words conceptually, i.e., within the context of their conceptual referents. The predominate associates were synonyms or superordinates of SWs. This tendency to classify or to define SW as a concept was particularly evident in the Multiword responses of Factor III subjects, e.g., PRESS, BEAR DOWN; CIRCULATE, MOVE ABOUT. Also, on SWs that usually evoke Contrast associates, Factor III subjects occasionally "passed up" a high-commonality Contrast response to give a Multiword definition, e.g., SHORT, NOT LONG. Vocabulary correlated .56 with the Synonym variable (Table 1, total 4 days) and probably was an important source of variability on Factor III. Low-Vocabulary subjects with a set to give conceptually related responses were a source of the Distant associates on this factor, e.g., RELISH, PLANT.

In many response tendencies Factor IV subjects were very similar to Factor III subjects (see Table 4). It would appear, however, that subjects high on Factor IV took the examiner's instructions literally to "Give the *first* word." The great frequency of Contrast and Coordinate associates, high Commonality, and fast Reaction Time that characterize Factor IV suggests the speed orientation of these subjects. The extreme speed with which Contrast and Coordinate associates can be given has been noted by several investigators (Karwoski & Schacter, 1948; Wells, 1911; Woodworth &

Wells, 1911).[4] Moreover, when the same subjects are retested under intense time pressure, the number of Contrast and Coordinate associates given more than doubles (Siipola, Walker, & Kolb, 1955). Despite the predominance of Contrast and Coordinate associates on Factor IV, however, it would be more accurate to characterize this set as one to give the *first* possible response, rather than as a set specifically to give contrasting associates. For example, on SWs that tend to elicit an extremely popular Functional or Superordinate primary response, the subjects high on Factor IV often gave the primary response more frequently than Factor II or III subjects. The 12 Factor IV subjects presented in Table 5 responded THREAD to NEEDLE more frequently than the 12 Factor II subjects, though they could have given a coordinate, PIN. Similarly, PARROT evoked BIRD more frequently from the Factor IV subjects than from the Factor III subjects, despite the "availability" of many coordinating associates. On the other hand, Factor IV subjects more frequently gave low-commonality, banal associates such as RULE, RULER; STAIR, WAY. Factor IV appears to represent a kind of linguistic-reflex, or speed, set resulting, incidentally, in many Contrast and Coordinate associates.

A more complete understanding of these three idiodynamic sets might be attained by means of controlled association experiments, in which systematic modifications of explicit instructions are used by the examiner to induce in the subject an object referent, a conceptual referent, or a speed orientation. The explicit instructions that evoke response patterns resembling those in Factors II–IV should supply clues to the implicit instructions under which subjects high on these factors were operating.

The influence of idiodynamic sets in facilitating or interfering with the mediation of SR word connections should not be overlooked in word-association studies. As shown in Table 5, the largest number of primary (and fewest individual) responses was given to SWs most compatible with the subject's set. On the other hand, it may be seen in the same table that a set of one kind actively interfered with mediation of responses to SWs that were more compatible with one of the other sets. To illustrate, the subjects with a conceptual-referent set would tend to achieve a higher Commonality score and commit fewer mediation faults in associating to a list consisting of concept words like ANGER, BEAUTIFUL, and WISH. The subjects with an object-referent set would, however, tend to have more mediation faults and lower Commonality in associating to these words. The situation would be reversed if the list consisted of "thing" words like STOVE, RIVER, and WINDOW. According to the Minnesota norms (Russell & Jenkins, 1954), these two lists of SWs have the same Commonality value. Clearly, word lists equated for experimental purposes by means of Commonality norms are not actually equated for all subjects. An equal number of SWs compatible with each of the different idiodynamic sets seems to constitute an additional requirement for equated word lists. For any SW list, the proportion of words that are compatible with the subject's idiodynamic set is a significant determinant of the Commonality score he achieves and of the number of mediation faults he commits.[5] Commonality itself is a complex variable, a resultant score derived from consensus within different set types.

CONCLUSIONS

1. The present study corroborates the observation of other investigators that, in the free-association experiment, individuals exhibit characteristic associative sets.

2. Three idiodynamic sets were identified:

---

[4] A Contrast associate tends to be the terminal link in a chain of association. If the subject were to associate next to his own response word, the most probable associate would be the original stimulus word, e.g., BLACK, WHITE; WHITE, BLACK. This is not as true of the other structural variables, e.g., BARK, TREE; TREE, LEAVES; OT, CABBAGE, VEGETABLE; VEGETABLE, FOOD. The paucity of competing associates may account in part for the faster reaction time of Contrast associates.

[5] The 100-word Kent-Rosanoff list, the most widely used list in the search for "correlates of Commonality," is particularly defective. Of the 100 primary (most popular) response words, 38 are Contrast or Coordinate associates, yielding a sum Commonality score of 1,746; only 10 primary response words are Synonym or Superordinate associates, yielding a sum Commonality score of 351 (Russell & Jenkins, 1954).

(*a*) Object-referent set—the subject interprets the task as one of naming an object associated in everyday experience with the object, attribute, or action denoted by SW. It was hypothesized that an important determinant of the subject's interpretation was a concrete, denotative attribute toward words in general. (*b*) Conceptual-referent set—the subject interprets the task as one of specifying a word that is related logically (in the sense of relationship to other concepts) to the word named by the examiner. An abstract, conceptual attitude toward words in general was hypothesized to account for the subject's interpretation of instructions. (*c*) Speed set—the subject interprets the task literally to respond as fast as possible. Other subjects generally can be induced to produce responses characteristic of this set by the simple expedient of emphasizing the importance of speed (Siipola et al., 1955).

3. The subjects consistently exhibited the same idiodynamic set on different occasions and in responding to different word lists.

4. Idiodynamic set facilitates high commonality associates to SWs that are most compatible with the subject's specific set, but interferes with association to words that are more compatible with another set. (*a*) The subjects tend to give the most popular response to SWs that are compatible with their set, but in remaining consistent within their set frequently give uncommon responses to words that are more compatible with other sets. Hence, the Commonality score the subject achieves is a partial function of the proportion of SWs in the list that happen to be compatible with his set. (*b*) Mediation faults occur most frequently on SWs that are incompatible with the subject's set. In clinical applications of the word-association technique, the subject's idiodynamic set should be considered in the interpretation of so-called "complex indicators." (*c*) It follows that word lists equated for grammatical form, familiarity, and commonality value still may vary appreciably in difficulty level, depending on the subject's set. Equated word lists should have an equal number of words compatible with at least the three idiodynamic sets so far identified.

5. On each of the 4 days, a major factor appeared, characterized as "general ability to cope with the task," which was independent of the set factors.

IDIODYNAMIC SET IN ACUTE SCHIZOPHRENICS

Perhaps the most striking examples of bizarre thinking, of associative disorganization, are to be found among schizophrenic patients. The word-association performance of schizophrenics is characterized by Distant associates, erratic reaction times, perseverations, neologisms, etc. (Rapaport, Gill, & Schafer, 1946). Most commonly, however, these aberrant responses occur amid a majority of normal responses. Do the normal associates given by schizophrenics follow a systematic pattern; and, if so, does this pattern resemble the idiodynamic sets identified in normal subjects? The present study addresses itself to these questions.

The 79 hospitalized schizophrenics examined in this study were tested, when circumstances permitted, while acutely psychotic, preferably shortly after admission to the hospital. The intent was to obtain the best possible record of clearly schizophrenic associations. It will be recalled that these schizophrenics were matched closely for age, education, and intelligence (Vocabulary) with the 79 normal subjects in the preceding study. The same word lists were administered to the schizophrenic subjects in exactly the same manner as to the normal subjects. Statistical procedures were identical to those described in the preceding study.

RESULTS

The means, standard deviations, and intercorrelations of the 15 variables on each day are provided in Table 6. On each of the 4 days the schizophrenic subjects scored lower than the normal subjects on all 6 structural variables. As expected, they committed substantially more of each of the five mediational faults on every day. Commonality scores of the schizophrenics on all days, it follows, were lower than those of the normal subjects. Clearly, the two samples of subjects must be considered as representing different populations, with respect to mean levels on these 12 response variables (compare Tables 1 and 6).

TABLE 6

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF 15 VARIABLES FOR 79 SCHIZOPHRENIC MEN ON 4 SUCCESSIVE DAYS

Variable	*M*	*SD*	Day 1 intercorrelations													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Synonym	13.5	6.9														
2. Contrast	8.8	6.8	12													
3. Coordinate	8.3	5.5	26	69												
4. Superordinate	10.9	6.3	49	—08	—03											
5. Subordinate	3.8	2.2	44	—03	11	35										
6. Functional	15.3	7.3	07	09	—05	18	16									
7. Multiword	14.1	20.8	—26	—49	—49	—24	—20	—32								
8. Blank	6.4	10.5	—22	—10	—30	01	—13	—13	06							
9. Distant	7.6	9.1	—34	—10	—04	—37	—36	—26	—23	—14						
10. Reaction Time (seconds)	3.7	1.4	—04	—41	—32	17	04	—03	15	23	—01					
11. Reproduction Failure	37.2	29.2	—47	—41	—40	—40	—35	—51	47	20	41	31				
12. Commonality (%)	9.7	3.6	55	61	52	51	27	44	—59	—28	—35	—33	—71			
13. Age	35.3	7.4	—03	—26	—30	08	07	—04	20	11	—01	01	07	—19		
14. Education	10.9	3.2	10	22	23	—03	13	05	—21	—31	02	—38	—25	26	—25	
15. Vocabulary	20.3	6.7	47	32	32	05	24	14	—21	—38	—16	—43	—39	45	—02	49
			Day 2 intercorrelations													
1. Synonym	9.4	6.2														
2. Contrast	8.3	6.7	—07													
3. Coordinate	8.2	5.2	13	71												
4. Superordinate	11.5	7.1	60	—08	—04											
5. Subordinate	2.6	1.5	31	08	12	16										
6. Functional	16.1	7.1	23	14	10	23	15									
7. Multiword	10.8	18.3	01	—43	—36	—09	—20	—41								
8. Blank	5.4	11.2	—21	00	—21	—07	—06	—10	—01							
9. Distant	7.3	8.9	—43	—15	—14	—39	—18	—45	—15	—14						
10. Reaction Time (seconds)	3.7	1.4	13	—43	—30	30	—00	01	06	10	07					
11. Reproduction Failure	35.4	30.1	—36	—34	—37	—25	—21	—49	35	11	51	14				
12. Commonality (%)	10.0	3.8	35	64	52	39	26	59	—54	—19	—44	—36	—57			
13. Age	35.3	7.4	04	—07	—11	09	15	04	13	04	—07	06	—06	—08		
14. Education	10.9	3.2	04	08	09	—07	09	09	—14	—19	—07	—22	—04	12	—25	
15. Vocabulary	20.3	6.7	41	29	30	19	22	19	—22	—34	—25	—36	—38	46	—02	49
			Day 3 intercorrelations													
1. Synonym	9.8	6.8														
2. Contrast	7.8	8.0	04													
3. Coordinate	9.1	5.2	28	70												
4. Superordinate	9.7	6.5	65	—11	14											
5. Subordinate	2.2	1.9	19	21	38	25										
6. Functional	13.5	6.3	12	10	25	28	28									
7. Multiword	11.4	17.7	—05	—34	—38	—03	—14	—29								
8. Blank	6.6	13.4	—16	—11	—24	—05	01	—02	—06							
9. Distant	7.5	10.8	—34	—09	—19	—38	—32	—35	—14	—13						
10. Reaction Time (seconds)	3.7	1.4	30	—38	—25	51	—11	03	08	09	—05					
11. Reproduction Failure	40.9	27.7	—27	—36	—48	—27	—23	—41	35	06	54	13				
12. Commonality (%)	8.9	3.5	39	63	73	41	47	59	—43	—15	—43	—22	—63			
13. Age	35.3	7.4	05	—06	04	17	—07	15	05	07	—02	10	—04	01		
14. Education	10.9	3.2	12	00	09	01	00	—07	—15	—21	—01	—22	—04	06	—25	
15. Vocabulary	20.3	6.7	52	36	47	16	12	06	—22	—29	—20	—25	—33	49	—02	49
			Day 4 intercorrelations													
1. Synonym	11.2	6.6														
2. Contrast	7.2	5.9	07													
3. Coordinate	9.6	5.9	26	69												
4. Superordinate	11.1	7.5	57	—25	—12											
5. Subordinate	3.6	2.4	49	23	35	32										
6. Functional	15.5	8.0	21	15	10	31	30									
7. Multiword	9.1	15.7	—14	—39	—27	—11	—05	—38								
8. Blank	6.3	13.3	—22	—16	—30	04	—31	—10	—03							
9. Distant	6.3	8.5	—32	—12	—10	—34	—33	—30	—08	—09						
10. Reaction Time (seconds)	3.5	1.5	33	—42	—23	41	—10	10	08	07	08					
11. Reproduction Failure	35.3	28.5	—40	—35	—39	—23	—27	—44	44	17	49	12				
12. Commonality (%)	10.0	3.8	45	61	47	34	57	67	—46	—22	—41	—19	—62			
13. Age	35.3	7.4	01	—12	—17	11	12	07	06	03	—12	05	—07	—03		
14. Education	10.9	3.2	17	13	18	—05	03	—08	—17	—17	—07	—22	—09	15	—25	
15. Vocabulary	20.3	6.7	46	40	39	11	32	05	—24	—31	—19	—27	—38	46	—02	49

The rotated factor loadings on each of the 4 successive days are shown in Table 7. Examination of this table quickly shows that results over the 4 days were quite variable and cannot be discussed collectively at the outset as in the preceding study of normal subjects. Initially, the results of Day 1 in Table 7 will be emphasized. Data from several other sources then will be marshaled prior to entering into the intricacies of results on the other 3 days.

Factors I–III in Table 7, Day 1, indicate that Functional, Contrast-Coordinate, and Synonym-Superordinate-Subordinate represented three essentially independent modes of responses. This is much the same as the results for normal subjects, with Subordinate responses entering more strongly on the conceptual-referent factor for the schizophrenics. Factor V appears to be an "ability to cope with the task" factor, with Education and Vocabulary high positive and reaction time and Blank high negative. In the schizophrenic sample Distant associates on Day 1 tended to be independent of the set factors (except Factor III) and to characterize a separate factor, IV. On the first day they were tested these schizophrenic subjects certainly may be said to have revealed idiodynamic sets closely resembling those found in normal subjects.

In the preceding study it was found instructive to examine the relationship of the structural variables to other response tendencies. An identical analysis for the present sample is presented in Table 8. A comparison of Table 4 (normal subjects) with Table 8

TABLE 7

ROTATED ORTHOGONAL FACTOR LOADINGS ON 4 SUCCESSIVE DAYS

Variable	I	II	III	IV	V	h^2
Day 1						
1. Synonym	05	−19	81	−12	22	76
2. Contrast	−12	−83	−11	−31	18	85
3. Coordinate	10	−83	10	−09	27	78
4. Superordinate	−25	−01	77	−04	−19	69
5. Subordinate	−07	07	70	−06	16	52
6. Functional	−93	05	08	−05	07	89
7. Multiword	44	67	−21	−32	−06	80
8. Blank	05	03	−16	−31	−70	61
9. Distant	13	−08	−44	77	09	81
10. Reaction Time (seconds)	08	26	27	38	−63	69
11. Reproduction Failure	56	35	−41	32	−27	78
12. Commonality (%)	−45	−57	46	−22	26	86
13. Age	−04	51	03	−24	−03	32
14. Education	−04	−15	02	07	72	54
15. Vocabulary	−03	−14	26	−21	76	71
% Variance (71)	11	18	17	9	16	
Day 2						
1. Synonym	−53	07	16	−64	−06	72
2. Contrast	04	−37	−81	−10	−04	81
3. Coordinate	−20	−32	−71	−08	−06	65
4. Superordinate	−30	−13	34	−64	−14	65
5. Subordinate	−47	−16	−05	−21	−27	37
6. Functional	04	−45	−02	−63	12	61
7. Multiword	09	89	21	06	−10	85
8. Blank	74	02	07	−12	−16	59
9. Distant	−13	−25	17	88	−01	88
10. Reaction Time (seconds)	02	−22	79	−04	−22	72
11. Reproduction Failure	11	25	34	65	05	61
12. Commonality (%)	−20	−46	−50	−58	06	85
13. Age	−15	23	−04	−07	−72	60
14. Education	−30	05	−14	−04	76	68
15. Vocabulary	−59	05	−39	−33	36	74
% Variance (68)	11	11	17	20	9	

Table 7—Continued

Variable	I	II	III	IV	V	h^2
Day 3						
1. Synonym	−47	−05	69	−28	06	78
2. Contrast	−46	34	−44	−42	−15	72
3. Coordinate	−57	35	−17	−53	−14	77
4. Superordinate	−12	02	82	−34	−09	82
5. Subordinate	03	03	−00	−63	09	41
6. Functional	13	32	18	−59	−17	53
7. Multiword	04	−92	01	17	−14	89
8. Blank	68	07	−02	−14	−00	49
9. Distant	−04	39	−22	79	−06	83
10. Reaction Time (seconds)	27	04	79	20	−16	76
11. Reproduction Failure	19	−19	−05	72	02	59
12. Commonality (%)	−39	36	04	−79	−06	90
13. Age	−11	−03	16	03	−76	62
14. Education	−38	06	04	05	74	71
15. Vocabulary	−76	08	10	−25	30	75
% Variance (70)	15	10	14	22	9	
Day 4						
1. Synonym	−43	−23	61	−30	32	80
2. Contrast	−36	−57	−55	−02	16	78
3. Coordinate	−58	−43	−32	04	22	67
4. Superordinate	−02	−22	77	−35	03	76
5. Subordinate	−64	−20	14	−44	−04	67
6. Functional	−04	−72	23	−18	−25	67
7. Multiword	−10	81	01	−23	−18	75
8. Blank	80	−03	−02	−13	−07	67
9. Distant	−07	20	−07	85	−08	78
10. Reaction Time (seconds)	05	06	81	24	−20	76
11. Reproduction Failure	15	65	−02	41	−14	63
12. Commonality (%)	−36	−78	02	−35	12	87
13. Age	−06	10	00	−44	−50	45
14. Education	−03	00	−06	−04	86	75
15. Vocabulary	−41	−20	−07	−29	64	70
% Variance (72)	14	20	14	12	12	

TABLE 8

CORRELATES OF THE STRUCTURAL VARIABLES

Other response variable	Structural variable				
	Functional	Synonym	Superordinate	Contrast	Coordinate
16. RW intrinsic predicate of SW (e.g., APPLE, RED)	24	15	34	−11	−11
17. RW extrinsic predicate of SW (e.g., APPLE, ROTTEN)	−03	−26	−20	−04	00
18. SW intrinsic predicate of RW (e.g., RED, APPLE)	54	−01	01	−19	−13
19. SW extrinsic predicate of RW (e.g., ROTTEN, APPLE)	45	−25	−10	−25	−25
20. Action of or upon RW (e.g., HIT, BALL)	62	−25	−13	00	06
21. Associated actions (e.g., RUN, LEAP)	27	42	34	37	39
22. Total noun responses	54	17	18	06	23
23. Total verb responses	−17	31	27	−13	−03
24. Total adjective responses	03	22	05	49	49

Note.—Variables 16–21 and the 5 structural variables represent mutually exclusive categories, i.e., no response received more than 1 of these 11 scores. Grammatical form of RW, Variables 22–24, was not scored for ambiguous forms, e.g., noun-verbs, noun-adjectives, etc. Pearson *r's* were computed for total performance (on 500 SWs) of the 79 subjects.

(schizophrenic subjects) reveals the great similarity between the two samples. Schizophrenic subjects giving many Functional associates tended also to name objects in response to adjectives and verbs (Variables 18–20) and to give many nouns in general (Variable 22). Schizophrenic subjects giving Synonym-Superordinate associates tended not to give object-referent responses to adjective and verb SWs (not as marked as in the case of normals), but to give somewhat more verb responses (Variables 21 and 23). As in the analysis of normal subjects, schizophrenic subjects who gave many Contrast-Coordinate associates more closely resembled Synonym-Superordinate subjects, except for a tendency to give many adjective responses (Variable 24).

The factors appearing on Day 1, together with the correlates of the structural variables, afford for the schizophrenic sample evidence for the same three hypothesized idiodynamic sets as those appearing in the normal sample. In the preceding study of normal subjects successful differential predictions, based upon the hypothesized sets, were made for Commonality, individual responses, and frequency of mediation faults (see Table 5). The same predictions were made for the present schizophrenic subjects. In this case, however, the subjects were selected for specific set (same procedure as used with normal subjects) by their factor scores on Day 1 only. The same 60 SWs used in Table 5 were used to obtain the results shown in Table 9. As in the case of normal subjects, schizophrenic subjects with any one of the three idiodynamic sets achieved a higher Commonality score, gave fewer individual responses, and committed fewer mediation faults on the 20 SWs most compatible with his set.

The evidence to this point supports the view that the three idiodynamic sets identified in normal subjects were operative also in acutely psychotic schizophrenics. But, returning to Table 7, it may be seen that the factor analyses on Days 2, 3, and 4 bear little resemblance to the factors consistently found with normal subjects or to the Day 1 factors found with these schizophrenic subjects. Before attempting to describe the complex changes that evolved with successive retesting, one might ask the critical question: Do these changes indicate that the schizophrenics "lost" their sets after Day 1? A straightforward answer to this question may be obtained by attempting to make the predictions shown in Table 9,

TABLE 9

EFFECT OF SET (DETERMINED BY DAY 1 SCORES) UPON COMMONALITY AND MEDIATION FAULTS ON ALL DAYS

Type of set	Type of SW		
	Functional	Synonym-Superordinate	Contrast-Coordinate
Average commonality (%) of responses			
II. Functional	14.2	10.8	9.6
III. Synonym-Superordinate	9.5	14.0	9.3
IV. Contrast-Coordinate	10.9	9.1	17.7
Average frequency of individual responses			
II. Functional	8.2	10.3	10.5
III. Synonym-Superordinate	10.3	8.5	10.7
IV. Contrast-Coordinate	10.5	10.8	7.1
Average frequency of mediation faults			
II. Functional	12.3	16.2	16.4
III. Synonym-Superordinate	19.8	18.7	19.9
IV. Contrast-Coordinate	13.5	15.4	10.8

Note.—Each type of set is represented by 12 subjects with factor score greater than .50 on one set factor and less than .50 on the other set factors ($N = 36$) on Day 1. Each type of SW is represented by 20 SWs that elicited predominately one kind of response (Functional, Synonym-Superordinate, Contrast-Coordinate) to the exclusion of the other two kinds ($N = 60$ words). The three groups of SWs were equated (from norms on 196 subjects) for sum frequency of primary responses, i.e., Commonality value. Commonality of responses is the percentage of times each subject's response was given by 196 subjects. Individual response is given by less than 3% of 196 subjects. Mediation faults are all faults (can exceed one per word, e.g., a delayed Reaction Time, Distant response, with Reproduction Failure = three faults).

but to do so for Days 2, 3, and 4 only, using Day 1 set as the predictor. As may be seen in Table 10 such predictions can be made successfully. It would appear that the sets were operative in these subjects after Day 1.

Further evidence of the consistency with which these sets were operative over the 4 days is provided in Table 11. The scores treated in Table 11 are each subject's day-by-day deviations (*z* scores) from his own mean. Every subject thus had 4 *z* scores on each of the 12 variables. With 79 subjects, this procedure yielded 316 scores on each variable. The variables, so constituted, were intercorrelated and factor analyzed. Factors I, II, and III in Table 11 indicate that Functional (negative to Multiword), Synonym-Superordinate-Subordinate, and Contrast-Coordinate were response modes which tended to vary concomitantly over time, independently of each other. Factor IV indicates that Reaction Time and Blanks tended to vary concomitantly over time, independent of other response tendencies. These factors represent temporal dimensions and could emerge only to the extent that the subjects were consistent over the 4 days with respect to the idiodynamic sets.

Now that it is established that the idiodynamic sets were, in fact, operative after Day 1, the question arises: Why is this consistency not reflected in the factor solutions on Days 2, 3, and 4? Speaking loosely, the factor solutions underwent a complete reversal over time (see Table 7). On Day 1 the independent set factors were dominated by the very high loadings of the structural variables; by Day 4 the factors were dominated by independent mediation fault variables with the structural variables represented by low loadings scattered over several different factors.

The first clue was provided by the deviations from the subject's own means, which were calculated for the factor analysis shown in Table 11. Examination of the *z* scores of the 12 subjects selected to represent each set (i.e., used in Tables 9 and 10) revealed that, on Day 1, average *z* scores on the variables identifying their set were positive. On Day 4 their *z* scores on the same variables were negative. On the structural variables not representative of their set the opposite was true. For example, the 12 object-referent subjects had, on Day 1, an average *z* score of .29 on Functional and an average *z* score of −.55 on the other five structural variables. On Day 4 these *z* scores were reversed, being −.01 and .44, respectively. The same tendencies were found in subjects with conceptual-referent and speed sets. This finding indicates that there was some loss of set upon repeated testing. However, since these *z* scores were deviations from the subjects' own means they provide no

TABLE 10

Effect of Set (Determined by Day 1 Scores) upon Commonality and Mediation Faults on Days 2, 3, and 4

Type of set	Type of SW		
	Functional	Synonym-Superordinate	Contrast-Coordinate
Average commonality (%) of responses			
II. Functional	13.4	7.5	11.4
III. Synonym-Superordinate	8.9	11.4	10.3
IV. Contrast-Coordinate	9.8	5.6	18.4
Average frequency of individual responses			
II. Functional	8.4	10.4	9.6
III. Synonym-Superordinate	10.8	8.4	10.0
IV. Contrast-Coordinate	10.4	11.8	7.2
Average frequency of mediation faults			
II. Functional	12.4	15.6	15.8
III. Synonym-Superordinate	20.4	18.8	18.0
IV. Contrast-Coordinate	14.8	16.6	11.8

Note.—Same subjects and same SWs (omitting words used on Day 1) as shown in Table 9.

TABLE 11

ROTATED ORTHOGONAL FACTOR LOADINGS OF COVARIANCE SCORES

Variable	I	II	III	IV	h^2
1. Synonym	01	61	00	−09	38
2. Contrast	37	−24	58	03	53
3. Coordinate	−03	07	82	−05	68
4. Superordinate	09	67	−10	−29	55
5. Subordinate	05	41	21	29	30
6. Functional	63	35	−06	18	56
7. Multiword	−69	19	−02	01	52
8. Blank	−30	06	−07	−45	30
9. Distant	31	−47	−43	−25	57
10. Reaction Time (seconds)	01	08	08	−86	75
11. Reproduction Failure	−38	−16	−33	−05	29
12. Commonality	56	47	32	23	69
% Variance (51)	14	14	12	11	

Note.—Scores used in this analysis were, for each variable, each day's deviation score from the subject's own mean for that variable; hence, $N = 79 \times 4$ or 316.

information on the proportion of each kind of associate he was giving on Day 4; that is, on the relative "strength" of set on Day 4. This information may be derived from the subjects' deviations each day from the mean of the total sample.

Examination of daily deviation scores from the group means confirmed the above finding. For example, on Day 1, the 12 object-referent subjects had an average z score of 1.47 on the Functional variable and an average z score of −.35 on the other five structural variables. On Day 4 these subjects had an average z of .72 on Functional and a z of −.10 on the other five variables. Note, however, that these subjects were, on Day 4, still evidencing an object-referent set. Analyses of the deviation-score patterns of the subjects with a conceptual-referent or speed set yielded similar results.

One source of the differences in daily factor solutions now becomes clear. Each of the three sets was operative on all days but became less pre-emptive on repeated testing. It should be added also that the greatest change took place on Day 2. Rather than a progressive "weakening" of set each day, the most typical tendency was for a sharp break to occur between Days 1 and 2 with less change thereafter.

Mediation-fault variables dominated Factors I–IV on Day 4: I, Blank, loading .80; II, Multiword, .81; III, Reaction Time, .81; IV, Distant, .85 (see Table 7, Day 4). Reproduction Failure loaded on both Factors II (.65) and IV (.41). Thus, by Day 4 mediation faults were essentially independent of each other for these schizophrenic subjects. By using factor scores derived from the Day 4 analysis, it was possible to select 10 subjects as representative of each factor, i.e., with factor score greater than .50 on one factor and less than .50 on each of the other three factors. The z scores (deviation from total sample mean) of these subjects were then examined for all 4 days. A pattern of changes comparable to that found on the structural variables was discovered. However, instead of converging over time, like the z scores on the structural variables, z scores on the mediation-fault variables diverged over time. For the 10 subjects high on Factor I (Day 4), the average z scores on the Blank variable went from 1.54 on Day 1 to 2.02 on Day 4. For the Factor II subjects, the Multiword variable went from 1.19 to 1.69. On the Reaction Time variable Factor III subjects went from .65 to 1.25; and on the Distant variable Factor IV subjects went from an average z score of .63 on Day 1 to 1.85 on Day 4. In each of these groups of subjects the incidence of mediation faults of all other kinds was lower on Day 1 and tended to decline over the 4 days. It is evident that these subjects, on Day 1, committed one type of fault more frequently than any other and, on repeated testing, persisted in committing even more of this one kind of mediation fault and still fewer of the other faults. Thus, a second source of the differences in factor solutions on different days is identified.

Initially, to examine the influences of the three idiodynamic sets, 12 schizophrenic subjects representing each set were selected by means of Day 1 factor scores. Above, 10 subjects were selected by the same procedure to represent each of the mediation fault factors but from Day 4 factor scores. About half of the subjects selected by Day 1 factor scores also were selected by Day 4 factor scores. Since the same subjects often appeared in both analyses, it was possible to observe directly a third major source of the change in factor solutions on different days. Of the 12 object-referent subjects on Day 1, 2 were high on the Distant factor, 2 on the Blank factor, and 1 on the Reaction Time factor on Day 4. The conceptual-referent subjects on Day 1 scattered on Day 4, with 5 on the Reaction Time factor, 2 on the Multiword factor, and 1 on the Distant factor. Of the 12 speed-set subjects on Day 1, 3 appeared high on the Distant factor, 2 on the Blank factor, and 1 on the Reaction Time factor on Day 4.

Three major changes in the schizophrenic sample on repeated testing were: (*a*) a tendency for idiodynamic sets to become less preemptive, (*b*) a tendency for one kind of mediation fault (the one most frequent on Day 1) to become more stereotyped, and (*c*) a tendency for the subjects with the same idiodynamic set on Day 1 to evidence stereotypy on different mediation faults on subsequent testing. All of these changes, it might be noted, had a negligible effect upon group means on different days; the changes were offsetting. Perhaps it ought to be re-emphasized also that idiodynamic sets were still operative on all 4 days, despite the very marked changes in factor solutions that were generated, in large part, by the three trends described above.

DISCUSSION

Deliberate effort was made in the present study to obtain subjects who manifest a most severe form of associative disturbance, acutely psychotic schizophrenics. The success of this effort was evidenced by the marked mean differences between the present sample and the sample of normal subjects, particularly on the mediation faults, or complex indicators. Despite these large mean differences, the same three idiodynamic sets that were found with normal subjects were found operative also in the schizophrenics. The conclusions regarding idiodynamic set in normal subjects, listed in the preceding study, apply without alteration to these schizophrenic subjects, though some additional conclusions may be warranted by the present study.

To the degree that these idiodynamic sets may be used to represent association structure, normal and schizophrenic subjects appeared to share a common associative structure. Bopp (1955), using a different association technique, arrived at the same conclusion. She had 40 normal subjects and 40 schizophrenic subjects rate 16 words on a 13-scale Osgood semantic differential. Contrary to her original hypothesis of difference between the groups in semantic frame of reference, she concluded the following:

> Factor analyses of semantic judgments give the same semantic factors for the two groups, and the same semantic differential scales were even used to represent these dimensions of meaning. Not only is the semantic factor structure the same, but within the semantic space there is no appreciable group difference between concept meaning. Hence, schizophrenic thinking does not represent mere chaotic thinking, and it does not appear to touch the basic semantic structure [p. 69].

Taken in combination with Bopp's results, the discovery in schizophrenics of idiodynamic sets comparable to those of normal subjects suggests a useful differentiation within the broad term "association disturbance" between association structure and association process, i.e., mediational efficiency at a particular time. The distinction may have potential usefulness for better understanding of the nature of association disturbance in schizophrenia. The interpretation of a deviate response as the product of aberrant association structure has implications very different from interpretation of the response as resulting from a temporary "inaccessibility" to an otherwise normal association structure. The former interpretation, for example, implies the necessity of psychological treatment (psychotherapy) of schizophrenics. The latter interpretation admits the possibility of totally effective somatic treatment, e.g., correction of a biochemical abnormality which has acted to keep the individual in a confused state, resulting immediately in a "normal" person. The remarkable "normalizing" effects of tranquilizing drugs, with some schizophrenics, is a case in point. The reverse effect, the temporary "psychosis" produced in normal subjects by LSD, is another. In both instances, normal and psychotic association processes emerge and vanish, though association structure evidently remains unchanged.

Another suggestive bit of evidence is provided by a longitudinal study of one individual (Moran, Mefferd, & Kimble, 1960b). A 20-item word-association test was administered to a schizophrenic on 246 consecutive days. Only 30 alternate forms were used; hence, on every thirty-first day, he responded to a form he had taken 1 month before. At an arbitrary point (100 days) all subsequent responses were scored as "new" (not given previously to SW), or "old" (given previously to SW). The individual received chlorpromazine and electroshock in the fifth and sixth months. Shortly after cessation of electroshock treatment, he evidenced a marked clinical improvement and the number of new associations soared from an average of 38–64%. This increase in new associations was accompanied by a marked increase in Commonality scores. In the eighth month his condition deteriorated, accompanied by an encroachment of the (same) old associations (34–49%) and lower Commonality scores (62–42). Clearly, in this schizophrenic, features of a normal association structure were latent and emerged during the temporary improvement which followed somatic treatment.

To return to idiodynamic sets in schizophrenics, it was noted that such sets appeared most clearly at the time of first testing and were less evident, though still present, on subsequent retesting. Comparable changes in set were not observed in the small samples of normal subjects, though similar tendencies might well be detected with larger samples. Howard (1962), for example, has demonstrated convincingly in a series of five rigorous experiments with normal subjects, that significant changes do occur between the first and subsequent performances of subjects tested repeatedly, and that the second (or later) performance affords superior psychometric differentiation. The marked changes in the present schizophrenic samples bear out Howard's conclusions.

With regard to the nature of these changes in set on successive days, little may be added to the descriptive account given in the section on results. There it was noted that schizophrenic subjects with a definite idiodynamic set on Day 1 tended on subsequent days to give fewer of the associates that characterized his set and more of the associates that were typical of the other two sets. Sets were operative on all 4 days, but were weaker after Day 1. The changes that occurred in subjects with one set were compensated by changes occurring in subjects with other sets, so that the total sample means of the structural variables remained about the same on all days.

One possible lead to a better understanding of these temporal shifts in set is afforded by the concomitant changes in Commonality scores. The considerable accuracy afforded by Commonality norms in predicting response faults (Laffal, 1955) and even specific RWs (Russell & Jenkins, 1954) is well known. However, predictions based upon knowledge of the subject's idiodynamic set frequently go counter to those based upon Commonality norms. For example, from the Minnesota Commonality norms alone it would be predicted that a Minnesota student probably would respond CHAIR ($p > .80$) rather than FURNITURE ($p < .05$) to the word TABLE. Prior knowledge of a subject's concept-referent set would alter these probabilities, e.g., CHAIR ($p < .80$) and FURNITURE ($p > .05$). In the present samples of object-referent and conceptual-referent schizophrenic subjects, as sets became less pre-emptive, average Commonality scores increased. Put another way, on repeated testing the responses of these subjects became more predictable from Commonality norms alone. In the normal subjects this was not so; sets were sustained, relatively unchanged over the 4 days, and Commonality scores were very stable over the same period. Thus, schizophrenics with an object-referent or conceptual-referent set showed a decreasing tendency to strain for associates (frequently uncommon ones) consistent with their set and tended to give, instead, associates more consistent with those given by the average individual.

Commonality scores of Speed-set subjects did not increase on repeated testing. It will be recalled that the set to give the first possible response resulted, from the beginning, in many associates of high Comomnality. At this point it might be well to draw a sharper dividing line between the so-called Speed set and the other two sets. The object-referent and conceptual-referent sets are hypothesized to represent different orientations toward words in general. No such hypothesis is offered for the Speed set. Object-referent and conceptual-referent subjects alike may be induced to produce associates characteristic of Speed-set subjects, by application of time pressure (Siipola et al., 1955).

Another marked difference between the normal and the schizophrenic subjects was the tendency of the latter to commit increasingly more of one particular kind of mediational fault and fewer of the other kinds. Again, little may be added to this descriptive account. Normal subjects were highly consistent on all days with respect to both structural and mediation fault variables, as shown in Table 3. Different schizophrenic subjects with the same set on Day 1, however, frequently appeared with high factor scores on several different fault factors on Day 4. Thus, in normal subjects one may predict from idiodynamic set, to some extent, the kinds of mediation faults he is most or least likely to commit; in schizophrenics types of idiodynamic set and type of mediation fault were less closely related.

It was the purpose of these studies to identify idiodynamic sets and to examine some of their implications for the study of association structure by means of word association. The phenomenon of idiodynamic set seems to rest upon solid empirical ground, replicable over occasions with the same normal subjects and even in acutely psychotic schizophrenics. The rationales offered to account for these sets are, of course, very tentative. The authors have confirmed, with object-referent and conceptual-referent college students, the very marked increase in Contrast-Coordinate associates under intense time pressure, as reported by Siipola, Walker, and Kolb (1955). The term, Speed set, seems appropriate for the factors characterized mainly by Contrast-Coordinate associates. The basic differences in attitude toward words, hypothesized to account for the object-referent and conceptual-referent sets, is more speculative. Should these constructs prove useful, these idiodynamic sets ought to be assessed quite easily, without recourse to elaborate factor analyses as in the present study. For example, a word list consisting of homonyms that allow for a clear object-referent versus conceptual-referent associate might serve to differentiate the types, e.g., SWs like ALTER, CUFF, CURB, DUCK, HIDE, LOAF, NOVEL, VICE, and WATCH. Once differentiated, it would be of interest to determine the extent to which these two association tendencies are related to other cognitive processes. For example, in recall of a randomized list of homonyms, together with unambiguous object-referent and unambiguous conceptual-referent words, would subjects with an object-referent set tend to "cluster" (Bousfield & Cohen, 1955) the homonyms with the unambiguous object-referent words (e.g., ALE, WINE, WHISKY), and the conceptual-referent subjects cluster the homonyms with the unambiguous conceptual-referent words (e.g., AIL, WHINE, COMPLAIN)?

An interesting bit of confirmatory evidence for the constructs of object-referent and conceptual-referent associative sets comes from Ogden and Richards (1948):

> The metaphorical aspects of the greater part of language, and the ease with which any word may be used metaphorically, further indicate the degree to which, especially for educated persons, words have gained contexts through other words. Very simple folk with small and concrete vocabularies do, on the other hand, in some degree approximate to the account given above (p. 211),[6] since the majority of their words have naturally been acquired in direct connection with experiences. Their language has throughout many of the characteristics of proper names [p. 214].

---

[6] Ogden and Richards (1948) stated:

So far we have reached the level of the understanding of simple names and statements, and a considerable range of reference can be recorded and communicated by this means alone. A symbol system of this simple type is adequate for simple referents or aggregates of simple referents, but it fails at once for complex referents, or groups of referents which have a structure more intricate than mere togetherness [p. 211].

This statement was discovered by the present authors after the above formulations concerning idiodynamic sets were achieved. It would appear that Ogden and Richards, approaching the problem of "meaning" in a very different manner, had arrived at a highly similar conception of differences in attitude toward words. This observation by Ogden and Richards adds not one whit to the empirical evidence in support of the rationales for the idiodynamic sets; but, for the present authors at least, it contributes substantially to the verisimilitude of the constructs. Ogden and Richards attributed differences in attitude toward words to differences in education level. In the present study these differences were in the direction predicted by Ogden and Richards in both the normal and the schizophrenic samples. However, the differences were not large (conceptual-referent subjects averaged 5 months more education and 3.3 more words correctly defined on the WB vocabulary list, than the object-referent subjects) and are not very helpful in accounting for idiodynamic associate set differences.

CONCLUSIONS

1. With respect to mean differences, the sample of normal subjects and the sample of schizophrenic subjects may be said to represent different populations. On each of the 4 days the schizophrenic subjects scored lower than the normal subjects on all six structural variables and higher on all five mediation fault variables; the schizophrenics also had a lower Commonality score on each of the 4 days.

2. Despite these mean differences between the two samples on all 12 response variables, the idiodynamic sets found in normal subjects were found also in acutely psychotic schizophrenics. In fact, the conclusions derived from the preceding study of normal subjects apply also to the present sample of schizophrenics.

3. Although the three idiodynamic sets were operative on all days in the schizophrenic samples, marked changes occurred after Day 1. (*a*) Sets became less pre-emptive on repeated testing. In object-referent and conceptual-referent subjects, weakening of set was accompanied by an increase in Commonality score. This follows, since to maintain consistency of idiodynamic set frequently requires the giving of low Commonality associates (see Tables 5, 9, and 10). (*b*) As the Speed set became less pre-emptive, Commonality scores did not increase. It was noted that the Speed set itself tends to result in high Commonality responses. The Speed set may be of a different order than the other two sets; its rationale does not seem to require any particular attitude toward words but rather an attitude toward test taking. (*c*) Schizophrenic subjects tended to give progressively more of one kind of mediation fault and fewer of other kinds. The fault that became stereotyped was predictable from Day 1 performance, since it tended to be the most frequent fault given on that day. However, different schizophrenic subjects with the same idiodynamic set on Day 1 tended to become stereotyped on different faults on subsequent days.

4. It was suggested that association disturbance may be viewed as stemming from an aberrant association structure or from temporary inaccessibility of a latent normal structure. The present findings and those of Bopp (1955) support the latter view of association disturbance in schizophrenia.

SUMMARY

Individual differences in characteristic associative set were investigated. Initially, 100 normal subjects and 96 schizophrenic subjects were examined. The subjects from the two samples were then matched by pair for age, education, and vocabulary yielding 79 matched pairs. All subjects were tested individually under free-association instructions. Different lists of 125 words each were administered on 4 successive days. Responses were scored for Synonym, Contrast, Coordinate, Superordinate, Subordinate, Functional, Multiword, Blank, Distant, average Reaction Time, Reproduction Failure, and Commonality. Age, Education, and Vocabulary were included also, yielding a total of 15 variables. Separate factor analyses (principal components, varimax rotation) were conducted for each day.

Essentially the same factors emerged on each separate day in the analysis of normal subjects. Three orthogonal factors, character-

ized by Functional, or Synonym-Superordinate, or Contrast-Coordinate associates, were interpreted as representing idiodynamic associative sets. The three sets were further interpreted as object-referent (concrete, denotative attitude toward words), conceptual-referent (abstract, conceptual attitude toward words), and speed set (task-taking attitude of giving the fastest possible response), respectively. The subjects with one of the above sets tended to give high Commonality associates to SWs most compatible with their set and to commit more mediation faults on the other SWs. This was interpreted as reflecting the facilitative and inhibitive effects of idiodynamic set.

The same three idiodynamic sets were found in the schizophrenic sample. However, after Day 1 the sets, though still operative, were less pre-emptive. It was hypothesized that schizophrenics do not differ from normals in association structure.

REFERENCES

Bopp, Joan. A quantitative semantic analysis of word association in schizophrenia. Unpublished doctoral dissertation, University of Michigan, 1955.

Bousfield, W. A., & Cohen, B. H. The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. *J. gen. Psychol.*, 1955, **52**, 83–95.

Goodenough, F. L. Semantic choice and personality structure. *Science*, 1946, **104**, 541–556.

Howard, K. I. Differentiation of individuals as a function of repeated testing. Paper read at Society for Multivariate Experimental Psychology, Chicago, 1962.

Karwoski, T. F., & Berthold, F., Jr. Psychological studies in semantics: II. Reliability of free association tests. *J. soc. Psychol.*, 1945, **22**, 87–102.

Karwoski, T. F., & Schacter, J. Psychological studies in semantics: III. Reaction times for similarity and difference. *J. soc. Psychol.*, 1948, **28**, 103–120.

Kent, G. H., & Rosanoff, A. J. A study of association in insanity. *Amer. J. Insan.*, 1910, **67**, 37–96, 317–390.

Laffal, J. Response faults in word association as a function of response entropy. *J. abnorm. soc. Psychol.*, 1955, **50**, 265–270.

Moran, L. J., Mefferd, R. B., Jr., & Kimble, J. P., Jr. The objective measurement of psychopathology in longitudinal studies. *Trans. Fifth Res. Conf. Coop. Chemother. Stud. Psychiat. Res. Approaches Ment. Illness.* 1960, **5**, 106–111. (a)

Moran, L. J., Mefferd, R. B., Jr., & Kimble, J. P., Jr., Standardization of psychometric and psychodiagnostic tests for daily measurements in psychopharmacological research. *Trans. Fourth Res. Conf. Chemother. Psychiat.* 1960, **4**, 135–138. (b)

Murphy, G. An experimental study of literary vs. scientific types. *Amer. J. Psychol.*, 1917, **28**, 238–262.

O'Connor, J. *Psychometrics.* Cambridge: Harvard Univer. Press, 1934.

Ogden, C. K., & Richards, I. A. *The meaning of meaning.* New York: Harcourt Brace, 1948.

Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book, 1946.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr 66216.

Siipola, E., Walker, W. N., & Kolb, D. Task attitudes in word association, projective and nonprojective. *J. Pers.*, 1955, **23**, 441–459.

Tendler, A. D. Associative tendencies in psychoneurotics. *Psychol. Clin.*, 1933, **22**, 108–116.

Wells, F. L. Some properties of the free association time. *Psychol. Rev.*, 1911, **18**, 1–23.

Wells, F. L. Association type and personality. *Psychol. Rev.*, 1919., **26**, 371–376.

Woodworth, R. S., & Wells, F. L. Association tests. *Psychol. Monogr.*, 1911, **13**(5, Whole No. 57).

(Received May 7, 1963)

Vol. 78, No. 3 Whole No. 580, 1964
Psychological Monographs: General and Applied

EXPERIMENTAL STUDIES OF ASSOCIATIVE VARIABLES IN ORIGINALITY[1]

IRVING MALTZMAN, MARIGOLD BELLONI, AND MARTIN FISHBEIN[2]
University of California, Los Angeles

A series of experiments was conducted to study associative variables that may facilitate problem-solving performance on problems having a specific correct solution. Several experiments were conducted with a group form of the two-string problem. It was found that prior verbal learning of relevant response sequences failed to affect problem-solving performance in the predicted manner. A more extensive series of experiments was conducted employing the Remote Associates Test (RAT) of originality. Facilitation of solutions to the items was found to vary with the extent to which prompting stimuli were associated with the problem solutions. Manipulation of associative-variables reliably influenced performance on RAT. Problem-solving performance on RAT, in contrast, was relatively resistant to changes in "extrinsic" variables.

PREVIOUS research at the University of California, Los Angeles, on training variables influencing originality employed problem materials for which there is no one correct solution (Maltzman, Bogartz & Breger, 1958; Maltzman, Simon, Raskin, & Licht, 1960). Uncommonness of the obtained responses was taken as the criterion of originality. With such tasks it was found that inducing subjects (*S*s) to emit uncommon associations in a modified word-association situation resulted in an increased disposition to emit uncommon responses in new stimulus situations.

The general purpose of the present series of experiments was to investigate associative variables influencing performance on problems having a unique solution, where only one uncommon response is correct, or at best a small number of responses meet the criterion. Most of the experiments to be reported here attempted to prompt or increase the probability of occurrence of the problem solution by presenting stimuli or evoking responses associated with the solution response. This approach therefore assumes that the principles operating in originality, and problem solving generally, are not radically different from the principles operating in simpler kinds of learning situations (Maltzman, 1960).

Originality appears to be particularly complex and inscrutable. But such behavior, or any other phenomenon, seems complex only as long as we do not know the relevant antecedent conditions and principles governing the behavior. One of the most important antecedent conditions of problem-solving behavior is the past history of the component responses constituting the final response chain called the problem solution. Unfortunately, the past history of the relevant behavior is usually unknown and uncontrolled when the typical problem-solving experiment is conducted with college students. One alternative is to teach *S*s an artificial language which will later be employed in problem solving (Moore & Anderson, 1954). In this manner apparent control over the past history of relevant responses may be obtained. An approach of this kind has several limitations. One of them, perhaps the least serious, is the practical one of the large amount of time needed to teach *S*s even a limited language. In addition, the associative strengths induced do not begin to approximate those found in common usage in a natural language.

---

[1] This work was supported by contract Nonr 233 (50) between the Office of Naval Research and the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the University of Illinois.

A second approach, the one employed here, is to vary the reaction potential of pre-existing responses or stimuli associated with the problem solution. Past research of this kind has been of two types. One avenue of study has been to prompt relatively specific responses relevant to an appropriate solution (Judson, Cofer, & Gelfand, 1956). A second research approach has attempted to produce relatively nonspecific prompting or transfer effects which facilitate a particular problem solution (Duncan, 1961; Flavell, Cooper, & Loiselle, 1958; Maltzman, Brooks, Bogartz, & Summers, 1958). The present study reports the results of studies manipulating both kinds of experimental variable. The first series of experiments reported here employed a group form (Guetzkow, 1951) of the Maier (1930) two-string problem in an attempt to replicate and extend the studies of Judson et al. (1956). The second and more extensive series of experiments used the Remote Associates Test (RAT) of creativity (Mednick, 1962) as the performance measure.

Experiments by Judson et al. (1956) employed the group form of the two-string problem as a performance measure in one of a series of experiments designed to study associative variables influencing problem solving. The *S*s provided as many written solutions as possible to the problem of grasping both strings simultaneously even though they appeared to be beyond arm's reach. The pendulum solution, attaching a weight to one of the strings and swinging it in pendulum fashion, was considered the correct solution. Prior to the presentation of the problem, and without knowledge that the problem would be administered, *S*s received rote-learning sessions. The principal experimental group received different lists of words to learn in serial order. One of the lists contained the relevant response sequence ROPE, SWING, PENDULUM. The purpose of this training was to establish a response chain, and by its repeated occurrence in the verbal-learning situation increase its probability of occurrence in the problem-solving situation. If such a response chain occurs in the problem situation and is initially evoked by sight of a picture of two strings, it should mediate a pendulum solution to the problem. The experimental results obtained by Judson et al. (1956) tend to support such an interpretation.

Their results suggest that it is possible to manipulate the initial response hierarchy in a non-problem-solving situation and by so doing to facilitate solution of a problem subsequently encountered. Such a procedure would permit the study under relatively standardized conditions of all the variables known to influence verbal learning in relation to response sequences evoked in subsequent problem-solving situations. Problem-solving performance could then be studied as a function of variables known to influence verbal learning under the relatively standardized conditions developed for the study of rote learning.

The first two experiments attempted to replicate the studies of Judson et al. (1956) preparatory to the investigation of the effects upon problem solving of additional verbal-learning variables. The negative results of these experiments led to the second series of experiments which employed a different problem task.

EXPERIMENT I

The first experiment attempted to repeat what seemed to be the relevant conditions of the experiment by Judson et al. (1956). However, their study was designed to produce a specific prompting effect, establishment of a relevant response sequence in verbal-learning sessions which would be evoked in the problem situation. A condition was therefore added in order to determine whether a nonspecific positive transfer effect could be produced. A method of inducing uncommon word associations or originality training was employed for this purpose (Maltzman et al., 1960).

Method

Subjects. The *S*s were 282 students drawn from introductory psychology classes, 162 women and 119 men.

Procedure. Three groups received verbal learning prior to presentation of the group form of the two-string problem. The learning list consisted of 10 words. For Group X_1 the first three words were ROPE, SWING, and PENDULUM. The remaining seven words were unrelated fillers. Presentation of the first three critical words in sequence was designed to facilitate the occurrence of the pendulum solution to the two-string problem.

Group X_2 received ROPE, TIE, PACKAGE as the first 3 critical words on their list. This response sequence was designed to interfere with the occurrence of the pendulum solution. A third group, C_1, served as a control condition, receiving 10 words, all presumably unrelated to the solutions for the two-string problem.

These three conditions are similar to conditions employed by Judson et al. (1956). Words presented to *S*s were obtained from the lists used by Judson et al. (1956). There were five oral presentations of the list to each group at a 4-second rate. The three groups were instructed prior to the start of verbal learning that there would be a subsequent retention test. The test was administered after presentation of the two-string problem. At the time of learning the word list, *S*s were not told that the two-string problem would be subsequently given.

Two additional groups were employed which did not receive verbal learning prior to the problem situation. Group X_3 was given a word-association task which has been found to increase the uncommonness of responses (Maltzman et al., 1960). The stimulus words were not associated in any obvious way with the pendulum solution for the two-string problem. This condition was introduced in order to determine whether the word-association originality-training procedure would facilitate the occurrence of a specific uncommon response, the pendulum solution.

The final Group, C_2, served as another control condition, receiving only the two-string problem without prior verbal learning or word-association training.

Each *S* was given a 2-page booklet which consisted of instructions, an illustration of the problem situation, and space for writing solutions. A photographic reproduction of the figure used by Guetzkow (1951) in his group form of the two-string problem was presented with instructions to write as many solutions as possible in 16 minutes. All *S*s were further instructed to indicate which of their responses they believed to be the best solution of the problem.

Results

The number of *S*s giving a pendulum solution was determined for each group. Additional performance measures obtained were the total number of solutions offered, the rank order of the "best" solution, rank order of the pendulum solution, and the number of words correctly recalled in the retention test given the verbal-learning groups.

Table 1 shows the frequency with which a pendulum solution occurred in the different groups.

TABLE 1

FREQUENCY OF PENDULUM SOLUTIONS BY MEN AND BY WOMEN IN EXPERIMENT I

Group	Men		Women	
	Yes	No	Yes	No
X_1	13	13	15	21
X_2	14	11	7	35
X_3	14	5	11	23
C_1	9	11	9	19
C_2	21	8	7	16

A chi-square test of the relative frequency of occurrence of the pendulum solution in the five groups was not significant ($\chi^2 = 7.27$, $df = 4$, $p > .05$). Since a sex difference is characteristically obtained with the two-string problem (Duncan, 1961; Judson et al., 1956) men and women were compared regardless of condition. As suggested by Table 1 the men obtained reliably more pendulum solutions than the women ($\chi^2 = 23.46$, $df = 1$, $p < .05$). A comparison of the relative frequency of pendulum solutions by the men in the various conditions yielded a chi square of 6.40 ($df = 4$, $p > .05$). A comparable analysis of the womens' performance yielded a chi square of 6.04 ($df = 4$, $p > .05$). Although men and women differed reliably in their number of pendulum solutions, the two groups did not differ significantly in their fluency or total number of solutions. The mean number of solutions of all kinds was 7.10 for men and 6.83 for women ($F < 1$, $df = 1/279$).

The apparent inconsistency in the performance of men and women on the two measures may be reconciled by examining the relationship between fluency and the rank of the pendulum solutions for men and for women. For men, there was a significant product-moment correlation between these measures ($r = .561$, $df = 46$, $p < .05$). For women, the correlation was not significant ($r = .119$, $df = 38$). The difference in the relationship between the rank of a pendulum solution and the total number of different solutions suggests that only the associative hierarchies of men are ordered in a fairly systematic fashion, with relatively uncommon solutions such as the pendulum solution being low in the hierarchy. For women there is no correspondence between the uncommonness of a solution and its position in the hierarchy. A similar finding has been reported by Staats (1957).

Ranks of the solutions rated as best by men and by women were not significantly different, although there was a tendency for women to select a solution which appeared earlier than for men. The mean rank for men was 3.50, while for women it was 3.26 ($F=2.50$, $df=1/275$). Mean ranks of the "best" solutions obtained under the different treatment conditions produced a significant F, 15.45 ($df = 4/272$). The error mean square was 1.166. Mean ranks for groups X_1, X_2, C_1, X_3, and C_2 were 4.16, 3.16, 3.68, 2.75, 3.00, respectively. Groups receiving verbal learning rated their best solution as lower in their associative hierarchies than groups without verbal learning.

Another unanticipated but related result pertains to the number of solutions obtained by the different groups. Groups receiving verbal learning had significantly fewer problem solutions, Groups X_1, X_2, C_1, versus X_3, C_2 ($F = 22.28, df = 1/279, p < .05$). There was no significant sex difference or interaction.

The results obtained with the fluency measure suggest that verbal learning has a nonspecific inhibitory effect upon problem-solving performance. A re-examination of the frequency of pendulum solutions in the combined verbal-learning versus no-verbal-learning groups supports this interpretation. Men without verbal learning, Groups X_3 and C_2, obtained reliably more pendulum solutions than men receiving verbal learning, Groups X_1, X_2, C_1 ($\chi^2 = 4.99$, $df = 1$, $p < .05$). Women in the combined-verbal-learning versus nonverbal groups did not differ reliably ($\chi^2 = 1.00$).

The previous finding that rank of the pendulum solution is correlated with fluency for men but not women may explain the above results that verbal learning produced a decrease in pendulum solutions among men but not among women. Since an uncommon response such as the pendulum solution is low in the response hierarchy of men, but is not in any systematic position in the hierarchy of women, reducing the total number of solutions would reduce the occurrence of pendulum solutions only for men. Presumably the uncommon responses that are low in the hierarchy are most susceptible to interference because of their relatively weak associative strength. A reduction in the total number of solutions given by women as a result of interference with uncommon responses in their hierarchies would have no consistent effect on their frequency of pendulum solutions, since the pendulum solution is not necessarily low in their response hierarchies or even present.

The apparent interfering effects on fluency found in the groups that received verbal learning suggest that the amount of interference may vary with the amount of retention in these groups. Another consideration required examination of the results of the retention test. Without retention of the three critical words by the experimental groups, a necessary condition for facilitating or interfering prompting effects on the pendulum solution could not be met. It was this consideration that led to the administration of the retention test following the problem-solving task. However, examination of the retention date indicated that almost all experimental *S*s recalled the three critical words. Inadequate retention of the critical words could not be responsible for the apparent absence of differential treatment effects.

The results of Experiment I imply that verbal-learning groups were subjected to sources of interference absent in the remaining conditions regardless of what other variables were present in the latter conditions. Thus Group X_3 which received word-association training and Group C_2 which received no prior experience were each significantly superior to the experimental and control groups, X_1, X_2, and C_1, which had different word lists to learn. It cannot be concluded from the superiority of X_3 as compared to C_1 that originality training was effective, since X_3 and C_2 did not differ reliably. These results suggest that the main effects obtained were due primarily to the presence or absence of verbal learning regardless of the type of list learned.

The results of Experiment I appear to be at variance with those reported by Judson et al. (1956) who found a tendency for facilitation or inhibition as compared to a control group, depending upon the type of word list used. They found no overall depression in the frequency of pendulum solutions by the experimental groups.

There are many conditions that differed between the present experiment and the one by Judson et al. (1956). One possibility is that the critical difference between the two experiments was the temporal order of the test for recall and the two-string problem. Judson et al. (1956) gave a retention test for the word lists before introduction of the two-string problem. In Experiment I the retention test was administered after the two-string problem. The *S*s in both experiments had been instructed that they would receive a subsequent test of retention.

The decrement in problem-solving performance in Experiment I may not have been caused by verbal learning per se, but by the delayed retention test which probably induced rehearsal of the word lists. An obvious test of this hypothesis would be to present different verbal-learning groups with the retention test before and after the problem-solving task. Presentation of the retention test before the two-string problem would correspond to the procedure of the Judson et al. (1956) experiment, while administration of the retention test after the two-string problem would be a replication of the procedure used in Experiment I. The next experiment was designed to make this comparison.

EXPERIMENT II

Method

Subjects. The *S*s were 75 students drawn from introductory psychology. There were 11 men and 14 women in each of the three conditions.

Procedure. Group X_1 received the same treatment as the corresponding group in Experiment I, the word list designed to facilitate pendulum solutions and a retention test following the two-string problem. Group X_2 received the same word list, but obtained the retention test before presentation of the two-string problem. Group C corresponded to C_2 in Experiment I. They did not receive the verbal-learning task. Other details of the experimental procedure were the same as in Experiment I.

Results and Discussion

TABLE 2

FREQUENCY OF PENDULUM SOLUTIONS BY MEN AND BY WOMEN IN EXPERIMENT II

Group	Men		Women	
	Yes	No	Yes	No
X_1	6	5	1	13
X_2	8	3	7	7
C	8	3	4	10

Table 2 shows the relative frequency of pendulum solutions in the three groups. Chi-square tests indicated that for the combined sexes the relative frequency of pendulum solutions in the three conditions approached significance ($\chi^2 = 4.04$, $df = 2$, $p < .10 > .05$). An analysis of the proportion of pendulum solutions obtained by the men in the three groups was not significant ($\chi^2 = 1.09$). The proportion of pendulum solutions obtained by women in the three groups did, however, differ significantly in the predicted direction ($\chi^2 = 6.32$, $p < .05$). Women in the experimental group receiving the retention test prior to the two-string problem obtained a higher proportion of pendulum solutions than Group C, while women in the experimental group receiving the retention test after the problem-solving task obtained a lower proportion of pendulum solutions than the control group.

An analysis of variance of the total number of solutions offered by the three treatment groups indicated that they did not differ reliably. Men gave significantly more solutions than women ($F = 6.85$, $df = 1/69$, $p < .05$). The interaction between treatments and sex was not significant.

The results of this experiment suggest that the interference effects obtained with verbal learning in Experiment I may be due to the occurrence of the retention test following the problem-solving task. A greater number of pendulum solutions occurred when the retention test was given before the problem-solving task, although the difference was reliable only for women in Experiment II. However, the total number of solutions did not differ in the groups as implied by the interference hypothesis.

Since the number of *S*s in Experiment II was relatively small, reducing the power of the chi-square test, and the results of the experiment were ambiguous, it was deemed advisable to repeat the conditions of this experiment.

A third experiment employing 227 *S*s was therefore conducted replicating Experiment II. None of the main effects was significant, thus failing to reproduce the results of Experiment II.

Results from these first three experiments were obviously discouraging. Our inability to reproduce the results reported by Judson et al. (1956) indicated the inadequacy of the two-string problem as a dependable problem-solving task. Variability in our results from one experiment to the next also corresponded to the difficulty Judson et al. found in reproducing the results of their first experiment in a reliable fashion. This lack of reliability was not due solely to the inherent lack of power of the chi-square test of the proportion of *S*s obtaining pendulum solutions, the principal performance measure. The fluency measure analyzable with parametric statistics did not fare much better in its consistency from one experiment to another.

Criticism of the two-string problem has repeatedly occurred (Duncan, 1959; Ray, 1955). It had been hoped, however, that the group form might prove a more adequate performance situation. These hopes were not realized, and as a consequence we discontinued our research with the two-string problem and turned to another test situation which provided a more dependable performance measure, as well as possessing a number of other important attributes not possessed by the two-string problem.

Obviously, weaknesses inherent in the performance measure used do not mitigate entirely the negative results. There was no evidence of specific or nonspecific transfer effects induced by the relevant experimental treatments.

EXPERIMENT III

In this and subsequent experiments an attempt was made to obtain an independent estimate of the associative history of the relevant responses constituting the solution to a problem by using norms of associative frequency. Obviously, not all problems are presently amenable to such an approach. At the very least, norms of associative strength must be available. Fortunately, there is a set of problem materials presumably measuring originality which approximates this desirable feature. The problem materials employed here and in subsequent experiments were taken from Mednick and King's (Mednick, 1962) RAT of creativity.

Twenty different problems from the test were employed in the experiment. Each problem or test item consisted of three words. The *S*'s task was to find a fourth word that was related to each of the three words constituting a given problem, for example, the words SIXTEEN, HEART, COOKIES. The answer to this item is SWEET—sweet sixteen, sweetheart, and cookies are sweet.

An important feature of the test for our purposes is that the solution to each problem, such as SWEET in the above item, is a stimulus word in the Kent-Rosanoff word-association list, and most of the words composing a problem are relatively uncommon associates of their respective solutions. Solution of the test problems requires the evocation of remote or uncommon associations. The ability to produce such uncommon responses is generally assumed to be a characteristic of originality (Maltzman, 1960; Mednick, 1962).

An apparent limitation of RAT is that the associative relationship between the word items and the solution is not in the normative stimulus-response order. Problem solutions, responses, are stimuli in the Kent-Rosanoff list. Since extensive norms are not available for the associative frequencies between the problem items as stimuli, and the solutions as responses, it was necessary to assume that the associative strengths reflected in the norms for the Kent-Rosanoff list are bidirectional (Storms, 1958).

Assuming that the associative strengths between the words constituting a problem and its solution are known, it should be possible to vary systematically the frequency of correct solutions by varying the number and/or strengths of associative prompts. The present experiment was concerned with the effects on problem solving of the number of associative prompts and knowledge of results or confirmation.

Method

Subjects. The *S*s were 160 students from introductory psychology classes, 20 in each of the eight different subgroups.

Procedure. For the first 10 problems half the *S*s received the correct answer, confirmation, immediately after they had their opportunity to solve a given problem. The remaining *S*s received no information concerning the correctness or incorrectness of their solutions. Following the first 10 problems, each of these two main groups was divided into four subgroups receiving zero, two, four, or six presentations of the dominant associate of a given solution prior to the presentation of the problem. The associate was presented as a stimulus in a word-association task.

The zero-prompts group received no experience with the association task. After a short rest period they received the last 10 problems in the same manner as before. The remaining groups received either two, four, or six presentations of the dominant associate of each problem solution as the stimulus in a word-association task. Thus in the case of the illustrative problem which had the word SWEET as a solution, the dominant associate and stimulus in the word-association task would be SOUR. Word-association practice with a relevant stimulus word occurred immediately before each successive problem. The *S*s received the usual word-association instructions to respond as quickly as possible with the first word that came to mind, but to give a different response to successive presentations of the same stimulus word. Stimulus words for the word-association prompts and the problems were read aloud by the experimenter (*E*). An interval of 1 minute was allowed for the solution of each problem. Five seconds were permitted for an association the first time a stimulus word was presented. An interval of 10 seconds was allowed for subsequent associations to the same stimulus word. The experimental *S*s were further instructed that the associative training was a method of making them think of different associations and that it might improve their problem-solving performance.

The experiment was run in group form with the groups varying in size from approximately 5 to 10 in number.

Results

The number of correct problem solutions obtained by each *S* on the first and the last 10 RAT problems was determined. A 4×2 analysis of variance of the number of correct solutions on the first list was conducted. The factors were confirmation versus nonconfirmation and the varying number of associative prompts which were to be administered before the test list problems. Analysis of the subgroups which were to receive the different prompts was conducted in order to determine whether the random assignment of *S*s to the different groups was effective.

Neither the main effects nor their interaction was significant for confirmation versus nonconfirmation ($F = 1.70$, $df = 1/152$) or for number of prompts ($F < 1$, $df = 3/152$).

Figure 1 indicates the results obtained from the last 10 problems, the test series. It is apparent that the number of correct problem solutions tends to increase with the number of associative prompts, the amount of word-association training. Confirmation consistently resulted in a larger number of correct solutions than nonconfirmation of *S*s' problem-solving efforts.

An analysis of variance of the test series indicated that the prompting and confirmation effects were significant ($F = 6.45$, $df = 3/152$, $p < .05$; $F = 10.08$, $df = 1/152$, $p < .05$, respectively). There was no reliable interaction. The error mean square was 4.91. A trend analysis indicated that the linear trend of correct solutions as a function of number of prompts was significant under the nonconfirmation and under the confirmation conditions ($F = 7.20$, $df = 1/76$, $p < .01$; $F = 5.05$, $df = 1/76$, $p < .05$, respectively).

In his associative interpretation of individual differences in originality, Mednick (1962) suggested that the slopes of the associative hierarchies would differ considerably for high and low originality *S*s. His hypothesis implies that low originality *S*s would initially emit common responses with the commonness decreasing as a negatively accelerated function of repetitions. High originality *S*s, in contrast, would display a relatively flat hierarchy. Their initial associations would already be relatively uncommon so that further decreases in the commonness of their associations would be relatively slight. Mednick (1962) also re-

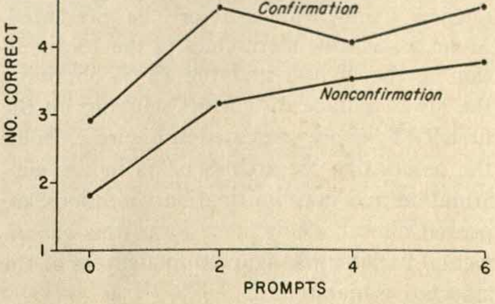


Fig. 1. Mean correct test solutions in Experiment III.

ports that performance on RAT is reliably correlated with associative originality.

The associative hierarchies obtained by *Ss* receiving six presentations of the word-association stimuli were examined in order to evaluate Mednick's hypothesis concerning individual differences in originality. Analyses of the associative hierarchies were limited to *Ss* in Group VI so that uncommonness of associations would be near asymptote.

Three different analyses were applied to the associative data. For all analyses the frequency of occurrence of the different responses to each stimulus word was determined for each presentation. In the first scoring method the rank of each association was determined and this rank was employed as the score assigned to each *S* emitting that association on the given presentation to the particular stimulus word. Each *S*'s associative hierarchy as a function of number of stimulus presentations was then obtained by determining the mean of the ranks of the associations to the 10 stimulus words at each repetition.

A second scoring method also employed ranks, but here the initial ranks were obtained for the frequency of a given association regardless of the particular stimulus presentation. Frequencies for each association to each stimulus were summed over all repetitions and then ranked.

A third scoring method used the obtained frequencies for each association on each repetition as the score assigned to each *S*'s responses. This procedure is the same as the first ranking method except that the raw frequencies were used rather than ranks.

Associative hierarchies obtained with the three scoring methods were highly similar in shape. Only the analysis employing the frequency scores will therefore be presented. Mean associative hierarchies of the 15 *Ss* obtaining the highest and the 15 *Ss* obtaining the lowest number of correct solutions on the first RAT list are presented in Figure 2. Since the associative hierarchies of *Ss* in the confirmation and nonconfirmation conditions appeared almost identical in form, this experimental variable was ignored in analyses of the associative hierarchies.

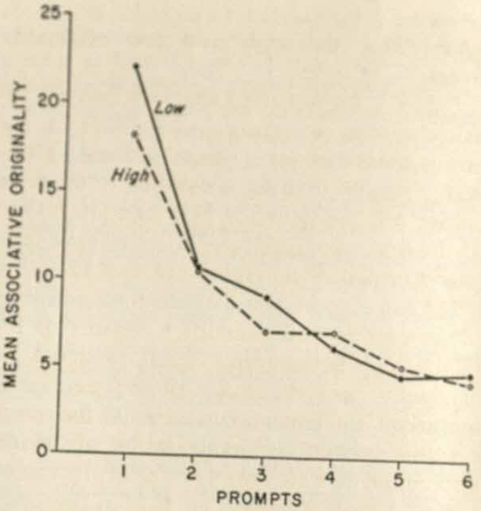


Fig. 2. Word-association hierarchies of high and low originality *Ss* in Experiment III.

It is apparent from Figure 2 that the associative hierarchies of high and low originality *Ss* are quite similar, and do not differ radically in shape as suggested by Mednick (1962). However, the results of an orthogonal polynomial analysis of trend indicated that the linear components of the two trends differed significantly, with the low originality group having the larger component ($F = 7.16$, $df = 1/28$, $p < .05$). This result would seem to support Mednick's hypothesis. However, inspection of the two curves indicates that the difference in the best fitting straight line for the two trends is determined by the first point in each curve. The overall shape of the trend for creative *Ss* is clearly different from that predicted by Mednick.

Quadratic components in the two curves did not differ reliably ($F = 1.39$, $df = 1/28$). Low originality *Ss* according to Mednick's hypothesis should have a reliably larger quadratic component, greater curvature, in the trend of their associative hierarchies.

It is possible that the limited range of scores provided by only the 10 problems of the first RAT list did not permit a sufficient differentiation of high and low originality *Ss*. Associative hierarchies were therefore analyzed from the highest and the lowest 15 *Ss* as determined by the total scores obtained from the initial and the final RAT lists. Neither the linear nor the quadratic components differed reliably for the high and low originality groups.

Figure 2 suggests that the associative originality of the two groups may differ on the first presentation of a stimulus word. However, a simple analysis of variance of the associative originality of the high and low groups on the first stimulus presentation did not indicate a reliable difference ($F = 2.19$, $df = 1/28$).

In order to obtain another measure of the relationship between individual differences in originality and the shape of the associative hierarchy, the linear components of the trends for the associative hierarchies of all 46 experimental *S*s were correlated with their number of correct solutions on RAT. The product-moment correlation between *L* scores and number of correct solutions on the initial RAT list was $-.14$ ($p < .05$). Since the restricted range of RAT scores would tend to suppress the size of the correlation, *L* scores for these *S*s were also correlated with the total number of correct RAT solutions on the initial and final tests. A correlation of $-.21$ was obtained between these scores ($p < .05$). Similar nonsignificant effects were obtained when the number of RAT solutions were correlated with the quadratic components of the trends of the associative hierarchies for these experimental *S*s.

Assuming that our method of obtaining uncommon associations meets his boundary conditions, the above results do not support Mednick's (1962) interpretation of individual differences in the creative process.

EXPERIMENT IV

Results from Experiment III showed that originality training, repeated evocation of associative responses to given stimulus words, facilitated the solution of verbal problems when the solution was associated with the stimulus word previously employed. Reliable facilitation was obtained in the absence of evidence that a direct prompting of problem solutions occurred. Problem-solving facilitation was obtained when solutions to the verbal problems were not evoked during associative training or were taken into consideration when they were evoked. An indirect transfer or prompting effect was thus obtained. However, the essential features of the prompting procedure responsible for facilitation of problem solving, or the extent of the nonspecificity of transfer, cannot be determined from Experiment III.

Additional research was therefore conducted in an attempt to determine more precisely the source of facilitation in the word-association task as well as to investigate other potential variables. In the present experiment one group of *S*s gave repeated associations to stimulus words that were the dominant associates of the problem solutions as in Experiment III. A second group gave repeated associations, emitted uncommon responses, to stimulus words that were not associated with the problem solution or the words constituting the problem. Facilitation of problem solving in this group would suggest that the originality training employed is relatively nonspecific. Prompting effects stemming from these stimuli and/or the evocation of uncommon responses are not restricted to the response hierarchy of which the associates are a member. A third group obtained prompts without emitting word associations. For these *S*s each problem item contained the dominant associative response to the solution in addition to the three standard words constituting a problem. The purpose of this condition was to determine whether it is necessary for the prompting effect to occur as a consequence of the emission of uncommon associations to a stimulus word, or whether a prompting effect may also be induced by the presentation of the stimulus word in the absence of associative responses. A fourth group received 10 additional problems of the same type prior to the final 10 test problems. This variation was introduced in order to determine whether the 10 additional problems would facilitate solutions to the final 10 problems, whether there would be any "learning to learn." The 10 additional problems also served as a control for fatigue and equated the total time spent in the experimental situation with the other experimental groups. The final group served as a control condition, receiving only the first 10 and the last 10 problems with no intervening experience.

Method

Subjects. The *S*s were 70 students drawn from introductory psychology classes, 7 men and 7 women in each of the five conditions.

Procedure. The *S*s were studied individually rather than in the group form used in Experiment III. They were all treated in the same fashion for the first 10 RAT problems. Each of the problems was presented individually on 3×5 inch index cards. A 2-minute time limit was enforced with a correction method. If *S* did not give a solution within the allotted time, or if he gave an incorrect solution, he was given the correct answer to the problem. If he gave the correct solution, he was told "right." Following these initial 10 problems, the five different treatments were introduced.

Group I was similar to the experimental groups of Experiment III, receiving originality training with the dominant associate of the solution to each of the last 10 problems.

Group II was treated in the same fashion except that the stimulus words used in the word-association task were unrelated to the words in the problems or their solutions as determined by examination of the Minnesota norms (Russell & Jenkins, 1954).

Group III did not receive word-association originality training. Instead, each of their last 10 problems consisted of four words. The first word in each problem was the dominant response in the associative hierarchy of the solution, the same words presented to Group I for originality training. The *S*s in Group III were instructed that their performance may improve as a result of having four-word problems. Group IV received 10 additional RAT problems between the initial and final 10 test problems.

Group C served as the basic control condition, receiving the final 10 test problems without any additional training or instructions.

Results

An analysis of variance of the number of correct solutions indicated that the groups and sexes did not differ significantly in the pretest phase, the first 10 problems. Since a fairly large within-groups regression coefficient was obtained between the initial and final test phases, .448, an analysis of covariance of the test phase was conducted. A significant treatment effect was obtained ($F=11.28$, $df=4/59$, $p<.05$). There was no reliable sex difference or interaction. The adjusted error mean square was 3.23.

Table 3 shows the results of individual comparisons of the group means based on *t* tests using the within-group variance as the error term. Group I, which received originality training with associated words, obtained significantly more solutions than all other groups except Group III, the prompting group which received four-word problems. The latter group obtained significantly more correct solutions than Group IV, the learning to learn group, and Group C.

TABLE 3

Adjusted Mean Correct Test Problem Solutions and Their Differences in Experiment IV

Adjusted *M*s		Differences between *M*s			
		II	III	IV	C
I	5.71	2.35*	.01	2.81*	3.44*
II	3.36		2.34*	.46	1.09
III	5.70			2.80*	3.43*
IV	2.90				.63
C	2.27				

* A mean difference of 1.53 is significant at the .05 level.

These results indicate that originality training, when it is effective in the present situation, must involve stimuli which evoke responses belonging to the same response hierarchy as the problem solution. Equally effective prompting of the solution may be obtained by introducing a dominant associate of the solution into the problem.

Errors due to failures to respond on the pretest and test lists were also analyzed. An analysis of variance indicated that the groups did not differ on the first 10 problems. Since the within-groups regression coefficient between the two lists was .718, an analysis of covariance of the test list was conducted. The results yielded a significant *F* for treatments and a significant interaction between treatments and sex (2.73 and 2.96, respectively, with $df = 4/59$ in each case). The adjusted treatment means and their individual comparisons are shown in Table 4.

These results differ in an important respect from the results obtained in terms of the number of correct solutions. In terms of failures to respond, the two groups receiving originality training, Groups I and II, did not differ significantly. Group II which had originality training with unrelated words obtained significantly fewer response failures than Group C.

Because of the highly skewed distribution of the failure to respond scores, Kruskal-

TABLE 4

ADJUSTED MEAN FAILURES TO RESPOND TO TEST PROBLEMS AND THEIR DIFFERENCES IN EXPERIMENT IV

Adjusted *M*s		Differences between *M*s			
		II	III	IV	C
I	.71	.65	.51	1.16*	1.79*
II	1.36		.14	.51	1.14*
III	1.22			.65	1.28*
IV	1.87				.63
C	2.50				

* A mean difference of 1.14 is significant at the .05 level.

Wallis and *U* tests were also conducted. Comparable results were obtained using these nonparametric statistics.

Results from the response failure measure suggest that originality training with unrelated words increases the disposition to respond as compared to Group C. However, the response that is uniquely correct is not differentially facilitated. Such facilitation, as manifested by Groups I and III, requires prompting from an associate.

An obvious shortcoming of this measure should be noted. The number of response failures is of necessity limited by the number of correct solutions. A significant difference in response failures between Group C and Groups I and III therefore was to be expected. Of greater theoretical import is the finding that Group II was significantly superior to Group C.

EXPERIMENT V

This experiment was designed to investigate further the implication of Experiment IV that a nonspecific effect of originality training is to decrease the number of failures to respond, to increase responsivity. A 3 × 2 factorial design was employed with originality-training groups that received associated and nonassociated stimulus materials and a control group. Two different instruction conditions were used in addition to the three training groups. Some *S*s received instructions to guess at solutions to RAT problems while others did not. If a nonspecific effect of repeated associative training is to increase responsivity, then a significant interaction should be obtained. The *S*s trained with unrelated stimulus materials should obtain reliably fewer response failures than control *S*s under the no-guessing, but not the guessing, instruction condition.

Method

Subjects. The *S*s were 75 students drawn from introductory psychology classes. There were 15 *S*s in each of the three training groups receiving no-guessing instructions and 10 *S*s in each of the three training groups receiving instructions to guess.

Procedure. Groups I, II, and C employed in this experiment received the same training as the corresponding groups in Experiment IV. Some of the *S*s within each of these groups received the same instructions as in previous experiments where nothing was said about guessing at solutions to RAT problems. For other *S*s the training conditions were repeated with one modification, guessing instructions were introduced for the last 10 problems. Fifty seconds after the presentation of each problem *E* instructed *S*s to guess if they had not obtained a solution to the given problem.

The experiment was conducted in group form with verbal presentation of the word-association and RAT items, similar in procedure to Experiment IV. Each RAT item was presented twice in succession. If a homonym appeared in an item, the word was spelled out on the first presentation. A 1-minute time limit was enforced for each problem. Time was recorded from the beginning of the second presentation of each item. At the end of the time limit *S*s were instructed to stop working on the problem and the correct solution to the problem was read to them.

Results

An analysis of variance of the performance on the first 10 problems showed that none of the *F* tests for treatments, sex, or their interaction was significant. This result held true for both the number of correct solutions and the number of failures to respond.

Table 5 shows the mean number of correct problem solutions and failures to respond in the test phase, the last 10 problems. Analysis of variance of the number of correct solu-

TABLE 5

MEAN NUMBER OF CORRECT (C) TEST PROBLEM SOLUTIONS AND RESPONSE FAILURES (F) IN EXPERIMENT V

Condition	Group I		Group II		Group C	
	C	F	C	F	C	F
No guessing	4.87	2.46	3.07	3.20	3.20	3.60
Guessing	5.70	.10	2.40	.60	2.60	1.1

tions resulted in a significant training effect ($F = 13.64$, $df = 2/69$, $p < .05$). Instructions and the interaction effect were not significant. The error mean square was 3.30. An analysis of variance of the number of response failures produced a significant instructions effect ($F = 28.59$, $df = 1/69$, $p < .05$). No other effects were significant. The error mean square was 3.19. Analyses of variance rather than covariance were employed because of the nonsignificant correlations between the first and last 10 problems.

Results of these analyses indicated that instructions to guess did not facilitate the number of correct problem solutions. The significant treatment effect on the number of correct solutions confirms the findings of Experiment IV. Despite changes in procedure introduced by group testing, comparable results were obtained in the two experiments. Originality training with the dominant associate of the solution resulted in significantly more correct solutions than training with an unrelated stimulus word and a control condition. The latter two groups did not differ significantly.

Results from the analysis of the response failure measure failed to show a significant treatment effect or the anticipated treatment by instructions interaction. Instructions did have a significant effect, with the groups instructed to guess giving reliably fewer response failures.

Thus this experiment failed to confirm the hypothesis that a nonspecific effect of originality training is to increase the disposition to emit responses. Although a slight difference appeared, Group II showing fewer response failures than Group C, the difference did not approach significance. There were also considerably more failures to respond in this experiment than in Experiment IV. The difference was probably due to the change in procedure introduced in the present experiment. Oral presentation of the materials to a group makes the problems more difficult.

EXPERIMENT VI

In the experiments previously described, it was found that associating repeatedly to stimulus words that were the dominant responses to the solutions of RAT problems produced a reliable increase in the frequency of correct solutions. Many questions remain unanswered concerning the limits under which this procedure is effective. The present experiment was designed in part to further explore the boundary conditions under which repeatedly responding to a relevant stimulus associated with the problem solution is effective. The previous experiments have shown that originality training with the dominant associate of the solution facilitates problem solutions, while originality training with an unrelated word does not. The purpose of the present experiment was to determine whether originality training with relatively weak associates from the response hierarchy of the solution facilitate performance or whether the prompting effect is limited to a small range of associates or perhaps only the dominant associate.

Another variation in the method of presenting the stimulus materials was employed in this experiment in order to explore the generality of the training procedure regardless of the particular mode of presentation. All materials were presented in booklet form for this purpose.

Method

Subjects. The *S*s were 180 students drawn from introductory psychology classes. There were 15 men and 15 women in each of the six conditions.

Procedure. Each of the items in RAT was presented on a separate page of a mimeographed booklet. The problem was repeated with the correct solution on the page following each of the stimulus items. The *S*s were instructed to work at their own speed; but they were further told not to take more than a minute or two for any one problem and that if they were not sure about an answer after about 2 minutes, they were to guess and go on to the next problem. They were not permitted to return to earlier problems.

Word-association originality training was given prior to each of the last 10 problems in the experimental conditions. Prior to the start of these problems *S*s in the experimental conditions were given additional written instructions concerning the word-association procedure. As in the previous experiments they were told that it was a method of making them think of different associations before each problem and that just thinking of many different associations in this way may improve their problem-solving performance. As in previous experiments it was emphasized that there was no necessary connection between the tasks so that they need not look for any hidden meanings.

Spaces for six associations to each stimulus word were provided in the booklet. Further instructions admonished *Ss* to respond each time to the stimulus word and not to their own associations and not to repeat any of their responses. As soon as *Ss* finished the first part of the experiment they went on to the test phase. All instructions and problem materials were in the mimeographed booklets.

Group X_1 received originality training with stimulus words that were the dominant associative responses to RAT solutions. Group X_{10} received originality training with stimulus words that were ranked tenth in the associative hierarchy of the solution words as determined by the Minnesota norms (Russell & Jenkins, 1954). Group X_{20} received stimulus words that were ranked twentieth in the hierarchy of the solution words. Group X_u was trained with words randomly selected from among the unique associative responses to the solution word. Group X_{un} was trained with unrelated words that did not appear in the associative hierarchy of the solution words. Group C received no originality training preceding the last 10 RAT problems. In the few cases where an association was a problem solution, the next ranking association was selected. Where more than one association had the same rank the tie was broken by random selection.

Results

An analysis of variance of performance on the first 10 problems indicated that the groups did not differ significantly prior to the introduction of the experimental treatments. Since there was a small but reliable within-groups regression between the pretest and posttest performance, .22, an analysis of covariance of the test performance was conducted. The adjusted mean correct test solutions for the different groups are shown in Figure 3.

An analysis of covariance of the number of correct test solutions indicated a significant treatments effect ($F = 2.38$, $df = 5/167$, $p < .05$). There was no significant sex difference or interaction between sex and treatments. The adjusted error mean square was 4.33. Individual mean comparisons employing the within groups variance as an error term indicated that Group X_1 receiving originality training with the dominant association obtained significantly more correct solutions than all other groups except Group X_{10}. None of the other groups differed significantly from each other. However, as Figure 3 indicates, there was a systematic ordering of the means in accord with the position of their stimulus words in the response hierarchy of the solution.

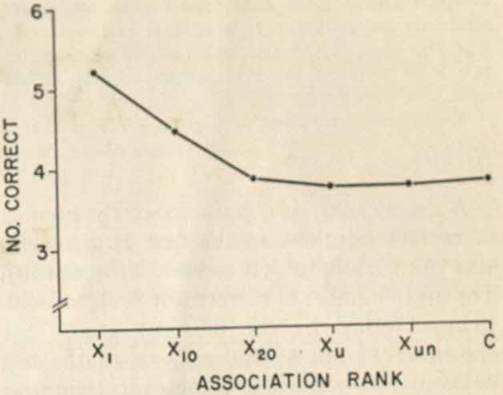


FIG. 3. Adjusted mean correct test solutions as a function of associative rank of prompts in Experiment VI.

The results of this experiment are somewhat ambiguous. Again, they demonstrate that associating to the dominant response in the hierarchy of the solution facilitates performance on RAT. However, Group X_1 did not differ significantly from Group X_{10} which associated to the tenth response in the associative hierarchy of the solution. But the latter condition did not differ significantly from the control condition. These results suggest that prompting effects produced by responding to associates of the solution are limited to a narrow range within the first 10 responses in the hierarchy, assuming that there is a reliable effect beyond the first response in the hierarchy.

EXPERIMENT VII

The ambiguous results obtained with Group X_{10} which did not differ reliably from Group X_1 or Group C, suggests that the associative hierarchy needs to be sampled within the limits of the first 10 associative responses. The present experiment was designed for this purpose.

Method

Subjects. The *Ss* were 136 students drawn from introductory psychology classes. The different groups contained from 21 to 25 *Ss*.

Procedure. The first, second, third, fourth, and fifth ranking responses in the associative hierarchies for each solution word were determined from the Minnesota norms. In cases where more than one associate had a given rank, one of the words was randomly selected for experimental use. In cases where a word in the association norms was an item in a RAT problem, the next ranking association was used.

Results

A simple analysis of variance of the number of correct solutions on the first 10 problems was not significant ($F = 1.99$, $df = 5/130$). The mean number of correct test-problem solutions obtained by the different groups are shown in Figure 4. An analysis of the test performance yielded a significant treatment effect ($F = 5.17$, $df = 5/130$, $p < .05$). The results of a Duncan range test indicated that all experimental conditions except X_4 differed significantly from Group C. Group X_1 also obtained reliably more solutions than Group X_4.

EXPERIMENT VIII

Experiment VII indicated that originality training facilitates solutions to RAT problems when the stimulus words are relatively strong associates of the solution. Results from Group III in Experiment II also indicated that the prompting effect may occur in the absence of evocation of the solution as a response in the word-association task. In the test series presented to this group the first word in each RAT problem was the dominant response in the associative hierarchy of the solution.

The purpose of the present experiment was to determine whether prompting effects may be obtained as a function of the associative ranking of the words in RAT items. Stimulus words used in Experiment VII for word-association training which were the first, second, third, fourth, or fifth associate of RAT solutions were added to each of the last 10 problems.

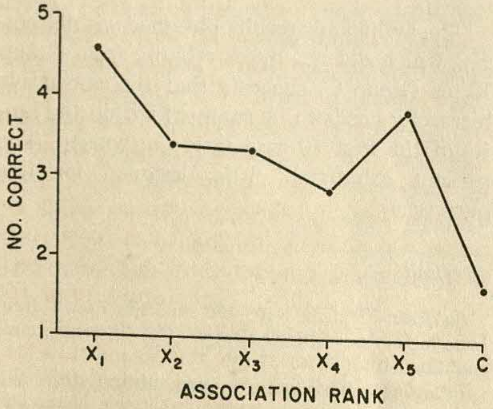


FIG. 4. Mean correct test solutions as a function of associative rank of prompts in Experiment VII.

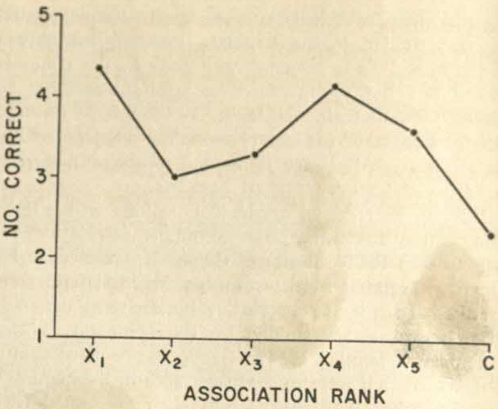


FIG. 5. Mean correct test solutions as a function of associative rank of prompts in Experiment VIII.

Method

Subjects. The *S*s were 120 students, 20 in each of the six conditions.

Procedure. Following the first 10 problems, different experimental groups received four words in each item constituting a RAT problem. The first, second, third, fourth, or fifth, ranking association to the solutions as determined from the Minnesota norms for the Kent-Rosanoff list were used as the first words in the appropriate items in RAT test. These are the same words used in Experiment VII for word association training. The usual control group was employed with the standard RAT items.

Results

A simple analysis of variance indicated that the groups did not differ in their performance on the first 10 problems. An analysis of variance of the test phase indicated that a significant treatment effect was obtained ($F = 3.42$, $df = 5/114$, $p < .05$).

The mean number of correct test problems obtained by the different groups are shown in Figure 5. A Duncan range test applied to these data showed that Groups X_1 and X_4 did not differ significantly. They each obtained significantly more solutions than Group C. Group X_1 was also significantly superior to Group X_2. There were no other significant differences among the groups.

Thus, in this experiment as well as in Experiment VII the frequency of correct solutions was reliably increased by associates of the solution beyond the dominant association in its hierarchy. However, the relationship is nonmonotonic, with the second most effective prompt produced by the fifth associate in Experiment VII and the fourth associate in Experiment VIII.

EXPERIMENT IX

In previous experiments employing associative training to facilitate performance on RAT, the influence of the stimulus word was not differentiable from that of the associative responses. One hypothesis is that the uncommonness of the associative response is primarily responsible for the facilitation of problem solutions. Occurrence of relatively uncommon responses prompts or increases the disposition to occur of other uncommon responses within the same hierarchy. An alternative interpretation is that the occurrence of the stimulus word prompts the solution response independently of the possible transfer effects stemming from the associative responses. Associative training may be effective for essentially the same reasons as the added associated words in RAT items of Experiment VIII. The alternative hypothesis is that when *S* is presented with a problem item, he recalls the stimulus word in the word-association series preceding the problem. Thus the situation becomes the same as when there are four items in a problem, except that now the fourth item must be recalled rather than responded to as part of a problem. Associating to the stimulus in the word-association originality training is thus unnecessary for the production of a facilitating effect.

A variety of experimental manipulations could be used to investigate the latter hypothesis. Only some of these possible variations were employed in the present experiment. They were designed to assess the relative contribution to the prompting effect of the stimulus words in contrast to associative responses.

Method

Subjects. The *Ss* were 180 students drawn from introductory psychology classes, 30 in each of the six different conditions. The experiment was conducted in group form similar to the procedure employed in Experiment III.

Procedure. The six groups were treated in the same fashion for the first 10 problems. Group C, the standard control condition, continued in the same manner for all 20 problems. Prior to each of the last 10 problems, Group I received a single presentation of the dominant associate of the solution response with instructions to give the first associate that came to mind. Group II received one presentation of the stimulus word with instructions to give six different associations to that word as quickly as possible. Group III was given six presentations of the same stimulus word but was instructed to write the first word they thought of only after all six stimulus presentations had been completed. Group IV received the standard experimental treatment of six presentations of the stimulus word with instructions to give a different association after each presentation. Group V was presented the six presentations of the same stimulus word with instructions to write down each stimulus word as soon as it was heard. There were no instructions to associate to the words.

Results and Discussion

Analysis of variance and the Duncan range test indicated that the groups did not differ in their frequency of correct responses on the first 10 RAT problems. A simple analysis of variance of the last 10 problems, the test series, indicated that the groups differed significantly ($F = 6.67$, $df = 5/174$, $p < .05$). The mean number of correct solutions and the results of a Duncan range test on the test performance are shown in Table 6.

Groups I and V tend to form one cluster while the remaining three groups form a second cluster. Each of the five experimental groups obtained significantly more problem solutions than the control, Group C. Group I, which received one stimulus word and gave one response, and Group V, which simply repeated the stimulus word each of its six presentations, did not differ reliably from each other, but did differ significantly from the remaining conditions.

At the very least, these results indicate that the overt occurrence of an associative response is not a necessary condition for prompting RAT solutions. They suggest further that

TABLE 6

DUNCAN RANGE TEST OF MEAN CORRECT TEST PROBLEM SOLUTIONS IN EXPERIMENT IX

Group	*M*	Group					Range
		IV	II	III	I	V	
		4.40	4.53	4.57	5.87	5.90	
C	3.20	1.20*	1.33*	1.37*	2.67*	2.70*	$R_2 = 1.092$
IV	4.40		.13	.17	1.47*	1.50*	$R_3 = 1.150$
II	4.43			.04	1.34*	1.37*	$R_4 = 1.189$
III	4.57				1.30*	1.33*	$R_5 = 1.218$
I	5.87					.03	$R_6 = 1.240$

* $p < .05$.

the principal prompting effect stems from the stimulus word rather than its associative responses. Conditions having more than one associative response obtain fewer problem solutions than the conditions where there are only one or no overt associative responses. An implication of the superiority of Groups I and V is that given the stimulus word, the occurrence of overt associative responses, may in fact interfere with facilitation of RAT solutions.

EXPERIMENT X

All of the preceding experiments with RAT have been concerned with variables that may influence problem-solving performance by changing the momentary reaction potential of the solution responses. The following experiments were concerned with the effects of variations in the experimental procedure, "extrinsic" variables, that presumably do not involve differential prompting of associative or solution responses. One such extrinsic variable is knowledge of results or confirmation, the effect on subsequent performance of informing *S* that he is correct or incorrect on a given problem. Experiment III indicated that knowledge of results affects RAT performance. The present experiment was designed to further explore this variable.

Method

Subjects. The *S*s were 80 students, 20 in each of the four conditions.

Procedure. A 2 × 2 experimental design was employed in which half the *S*s received confirmation of each solution for the first 10 RAT problems and half did not. Each of these two groups was subdivided before the last 10 problems so that half of the original confirmation group continued to obtain confirmation while half did not, and half the *S*s in the original nonconfirmation group continued as before while half now received confirmation for each of their last 10 problems. The remaining details of the procedure were the same as in Experiment III.

Results

A 2 × 2 analysis of variance was performed on the subgroups that were to receive the differential treatments in the test situation. Neither the *F* tests for the main effects nor their interactions approached significance.

A 2 × 2 analysis of variance of the number of correct solutions in the test phase indicated that neither main effects, correction versus noncorrection in the pretest, nor the test phase was significant. However, their interaction was significant ($F = 4.77$, $df = 1/76$, $p < .05$). Switching from one procedure in the pretest situation to another in the test resulted in superior performance than receiving the same procedure throughout both phases of the experiment. Receiving noncorrection in the pretest and correction in the test phase was superior to correction in the training and the test. Correction in the pretest and noncorrection in the test was superior to noncorrection in both phases. Noncorrection in both phases did not differ reliably from correction in both phases.

EXPERIMENT XI

Recent studies of paired-associate learning have found that Ss viewing paired associates learn at least as well as *S*s who actively anticipate the associates. An analogous problem has persisted in the area of problem solving in terms of guidance versus discovery of the solutions to problems. The RAT materials are readily adapted to the study of this problem by presenting the individual problems along with the solutions, guidance, in contrast to the usual procedure where *S* must discover the problem solution.

The present experiment was designed to investigate several variations and combinations of the effects of guidance, discovery, and knowledge of results on RAT performance.

Method

Subjects. The *S*s were 160 students drawn from introductory psychology classes, 20 in each of eight subgroups.

Procedure. The experiment employed a 2×2×2 factorial design. Half the *S*s received guidance and half received discovery conditions during the pretest training phase. Within each of these two conditions half the *S*s were asked to describe the relationship between the problem items and the solution to the problem. These are designated G_{II}, guidance, and D_{II}, discovery. Group G_{I} and D_{I} were not instructed to describe the relationship between the solution and the problem. For the test phase each subgroup was further subdivided, half the *S*s receiving knowledge of results or correction and half receiving noncorrection. The remaining aspects of the procedure were the same as in Experiment III.

Results and Discussion

A 2×2×2 factorial analysis was performed on the pretest scores of the groups as they would be formed for the test phase. None of the *F* tests for the treatments or their interactions was significant. A 2 × 2 × 2 factorial analysis of variance of the test scores likewise failed to show any significant treatment effects or interaction. Table 7 shows the mean number of correct test solutions in ascending order. A Duncan range test failed to show any significant differences. It is worth noting that the groups switched from correction to noncorrection were superior to the groups receiving correction in both phases of the experiment.

TABLE 7

MEAN CORRECT TEST SOLUTIONS IN EXPERIMENT XI

Condition	*M*	Range
G_{I}-NC	2.80	$R_2 = 1.36$
G_{I}-C	2.85	$R_3 = 1.43$
G_{II}-NC	2.95	$R_4 = 1.47$
D_{I}-G	3.00	$R_5 = 1.50$
D_{II}-C	3.33	$R_6 = 1.53$
D_{I}-NC	3.45	$R_7 = 1.54$
D_{II}-NC	3.55	$R_8 = 1.56$
G_{II}-C	3.90	

EXPERIMENT XII

This study repeated in part the conditions of Experiment XI because the results of the guidance treatments were somewhat ambiguous. They were neither reliably superior nor inferior to the conditions under which *S*s discovered the solutions to the initial problems. Actively solving the first 10 problems did not facilitate performance as compared to the guidance conditions.

Method

Subjects. The *S*s were 192 college students with from 30 to 36 *S*s in each of the six groups.

Procedure. Two of the six groups received nonconfirmation test phases. One of these groups received an initial series of problems with nonconfirmation while the other received no initial phase. Four groups received confirmation during their test phase. One of these groups also received confirmation during the initial phase. A second group received nonconfirmation initially, while two groups received prompting during the initial phase. One group obtained guidance with instructions to describe the relationship between the solution and problem words while the second guidance group received no special instructions. The remaining aspects of the procedure were the same as in Experiment III.

Results

A simple analysis of variance of the groups receiving RAT pretest indicated that they did not differ reliably prior to the experimental treatment ($F = 1.25$, $df = 2/91$). A simple analysis of variance of RAT test scores indicated that the various groups did not differ significantly ($F = 1.19$, $df = 5/174$).

Inspection of the means and a Duncan range test, although not significant, suggested that the guidance groups tended to be superior to the other conditions.

DISCUSSION

The present series of experiments was initiated in an attempt to investigate some of the variables that may facilitate performance in problem situations which have a specific response as a correct solution. Because such problems are assumed to require a relatively uncommon response on the part of *S* for solution, they may be said to require some degree of originality. Earlier research with so-called open-ended problem situations where there is no one correct solution (Maltzman et al., 1960) indicated that uncommon responses may be prompted by the repeated evocation of different associations to the same word-association stimuli. Not only did the successive responses to the same stimuli become more uncommon, there was an increase in the uncommonness of responses to new stimuli which had not previously been presented. The present series of experiments attempted to determine whether analogous effects could be obtained with problems which require a specific response as a solution. Under the conditions employed in the current experiments it is clear that such an effect was not obtained. Whether other kinds of training procedures or more extensive training regimes might prove effective are matters for further experimental research.

Results from the first three experiments employing the two-string problem failed to provide any evidence related to the generality of the training effect. Experimental effects produced by the verbal learning of relevant response sequences previously reported by Judson et al. (1956) could not be reproduced. Experimental effects obtained with RAT of originality appeared to be considerably more reliable. The subsequent discussion is therefore concerned only with RAT experiments.

Results from experimental conditions employed in Experiments IV and V demonstrated that under the conditions employed a nonspecific training effect cannot be obtained. Solutions to RAT problems were not facilitated when stimulus words were presented which were not associated with the solutions to RAT problems. When the word-association stimuli were associated with the solutions a prompting effect was obtained.

Successful solutions to RAT can be varied systematically by varying the number of presentations of the stimulus words in associative training as indicated in Experiment III. Facilitation may also be varied systematically by introducing stimulus words in associative training that vary in the assumed strength of their association with the solution responses. These results indicate that performance on this rather difficult test can readily be manipulated given some knowledge of the associative hierarchy of which the solution is a part. The apparent ease with which reliable facilitation can be obtained by associative prompting is especially striking when compared with the relative absence of reliable effects when extrinsic variables such as knowledge of results were varied. These results suggest that if information were available concerning the associative "pedigree" of solutions to various problems, performance on such tasks could be manipulated as readily as on RAT.

For purposes of the present series of experiments it was assumed that the association strengths represented by the Minnesota norms are bidirectional. If a stimulus word has associative strength for the evocation of a response, that response when serving as a stimulus word will have a similar tendency to evoke the former stimulus word as a response (Storms, 1958). In the absence of new norms an assumption of this kind is necessary in dealing with RAT because the solution response to each item is a stimulus word and the words constituting the problem are associations in the present norms. Their respective stimulus-response functions are reversed in RAT. Reliable effects in the present experiments produced by presenting associates of the solution as stimulus words in problems or in the word-association tasks support the bidirectionality assumption.

The results of Experiment IX cast considerable doubt upon the hypothesis that RAT solutions are prompted by the occurrence of associative responses that are members of the same response hierarchy as the solution. The two groups that were reliably superior to all others, Groups I and V, were the experimental groups that were required to give a minimum number of overt associative responses, one and none, respectively. They each obtained significantly more RAT solutions than Groups II and III which gave one and six responses to six presentations of the stimulus words. The pattern of results obtained by these four groups indicates that a greater number of overt associative responses is accompanied by a decrease rather than an increase in RAT solutions. Presentation of the stimulus word in the associative task is the critical variable, since all groups which received a stimulus word were significantly superior to the control group regardless of the responses evoked.

An increase in the number of RAT solutions in Experiment III as a function of increasing the number of prompts was not due to the evocation of progressively more uncommon responses, but the increased probability that the stimulus word would prompt the solution. Prompting of the solution response occurs as the result of mediated generalization, but not in the manner hypothesized initially. More likely, the stimulus word is recalled at the time the problem is presented. This implicit verbal response prompts the solution, in conjunction with the other words constituting the problem. The more often the stimulus word has been previously presented, the greater the probability that it will be recalled and prompt the solution. The stronger the association between the problem solution and the stimulus word as indicated by the word-association norms, the greater the probability that recall of the stimulus will prompt a solution when RAT is presented.

In Experiment IX Groups I and V, which gave only one and no responses, respectively, were superior in their RAT performance to that of the remaining experimental groups because they obtained a minimum of interference with recall of the stimulus word during presentation of a RAT item. A procedure such as the one used with the standard experimental conditions, according to the present interpretation, induces opposed tendencies. Repeated presentation of the stimulus word increases the probability of prompting the solution response, if the two are appreciably associated as indicated by the word-association norms. But the different responses evoked by the stimulus word and instructions by *E* interfere with recall of the stimulus word during presentation of RAT problem. Eliminating or reducing the number of associative responses, as in Groups I and V in Experiment IX, thereby results in an increase in problem solutions.

Precisely how the associative responses produce interference in recall of the stimulus words cannot be specified by the conditions employed in the present experiments. The assumed interference may be due to the usual form of retroactive inhibition. An alternative possibility is that the task of emitting associative responses induced *S*s to implicitly verbalize and this additional activity reduced the prompting effect.

Although the hypothesis that the associative responses prompt a solution does not appear tendable, the basic hypothesis that knowledge of the associative history of problem solutions permits their experimental control is supported by the results of the relevant RAT experiments.

In previous experimental studies employing open-ended problems where the responses were scored on the basis of their uncommonness with no one solution adopted as correct, it was found that associative training facilitated the occurrence of uncommon responses. Results from the present RAT experiments and their interpretation do not contradict these earlier results or their interpretation (Maltzman, 1960; Maltzman et al., 1960). The present experimental findings do indicate that the results in previous experiments cannot be generalized to RAT, and probably other problems with a specific correct solution, at least when the present amounts of training are employed. Effects of systematic training for weeks, months, much less years, remain to be investigated. In view of the very limited range of antecedent conditions explored in studies of originality training and the absence of established laws of thinking it is surprising to find the assertion that Maltzman's training procedure "may enrich one's fantasy life but without benefit to ingenuity of output [Caron, Unger, & Parloff, 1963, p. 442]." In order to substantiate such a statement, the laws of fantasy life as well as problem solving must be known. This obviously is not the case at the present time. It is our belief that the most effective way to obtain such laws is by stimulus-response analyses of the relevant associative and nonassociative variables and their systematic manipulation.

SUMMARY

A series of experiments was conducted to study associative variables that may facilitate performance on problems having a specific correct solution. Several experiments were conducted with a group form of the two-string problem. In addition to a control group, two different experimental groups learned word lists containing words designed to facilitate or interfere with solutions to the two-string problem. A third group received word-association training designed to increase the tendency to emit uncommon responses, thereby inducing a nonspecific positive transfer effect. This experiment and its two variations failed to provide reliable evidence of specific or nonspecific positive transfer effects to solutions of the two-string problem.

A more extensive series of experiments was conducted employing the Remote Associates Test (RAT) of originality. This test consists of a series of problems each consisting of three words. The task is to find a fourth word that is related to each of the three words constituting a given problem. Solutions to the problem items are stimulus words in the Kent-Rosanoff word list, and most of the words composing a problem are relatively uncommon associates of their respective solutions.

The first experiment in this series provided zero, two, four, or six presentations of the dominant associate to each solution in a free-association situation. Half the *S*s received confirmation after each problem solution and half did not. Frequency of correct solutions was reliably greater for confirmation than nonconfirmation groups, and increased reliably as a function of the number of associative prompts.

A second experiment was designed to further explore the variables in the prompting situation responsible for facilitation of problem solutions. In addition to two control groups, one experimental group associated to stimulus words that were the dominant associates of the problem solutions, as in the previous experiment. A second group received prompts from stimulus words not associated with the problem words or their solutions. For a third group each problem item contained the dominant associative response to the solution in addition to the three standard words constituting a problem. Only the first and third experimental groups obtained reliably more problem solutions than the control groups.

Additional experiments were conducted in order to determine whether prompting stimuli other than the dominant associate of the problem solutions would facilitate problem solving. It was found that stimulus words which were the first, second, third, and fifth ranking associates of the problem solutions reliably facilitated performance. Facilitating effects for the first and fourth ranking associates were obtained when the prompting words were presented as part of the problem items.

Another experiment was conducted in order to determine whether facilitation of problem solutions was produced by the prompting stimulus in the association task preceding each problem or the associative responses to the prompts. The results suggested that the stimuli rather than their associative responses were responsible for the facilitation of problem solutions.

Three experiments were concerned with the effects of "extrinsic" variables on problem-solving performance. Variations in knowledge of results and guidance versus discovery of solutions had relatively slight effects on the frequency of correct problem solutions.

Results from the present series of experiments indicate that performance on the problems employed which have only one correct solution can not be facilitated by the originality-training procedures successfully employed with tasks having no one correct solution. Whether variations of the present training procedure and more extensive periods of training would yield nonspecific positive transfer to problems possessing only one correct solution are questions requiring further experimental study.

REFERENCES

CARON, J., UNGER, S. M., & PARLOFF, M. B. A test of Maltzman's theory of originality training. *J. verbal Learn. verbal Behav.*, 1963, **1**, 436–442.

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

DUNCAN, C. P. Attempts to influence performance on an insight problem. *Psychol. Rep.*, 1961, **9**, 35–42.

FLAVELL, J. H., COOPER, A., & LOISELLE, R. H. Effect of number of preutilization functions on functional fixedness in problem solving. *Psychol. Rep.*, 1958, **4**, 343–350.

GUETZKOW, H. An analysis of the operation of set in problem solving behavior. *J. gen. Psychol.*, 1951, **45**, 219–244.

JUDSON, A. J., COFER, C. N., & GELFAND, S. Reasoning as an associative process: II. "Direction" in problem solving as a function of prior reinforcement of relevant responses. *Psychol. Rep.*, 1956, **2**, 501–507.

MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.

MALTZMAN, I. On the training of originality. *Psychol. Rev.*, 1960, **67**, 229–242.

MALTZMAN, I., BOGARTZ, W., & BREGER, L. A procedure for increasing word association originality and its transfer effects. *J. exp. Psychol.*, 1958, **56**, 392–398.

MALTZMAN, I., BROOKS, L. O., BOGARTZ, W., & SUMMERS, S. S. The facilitation of problem solving by prior exposure to uncommon responses. *J. exp. Psychol.*, 1958, **56**, 399–406.

MALTZMAN, I., SIMON, S., RASKIN, D., & LICHT, L. Experimental studies in the training of originality. *Psychol. Monogr.*, 1960, **74**(6, Whole No. 493).

MEDNICK, S. A. The associative basis of the creative process. *Psychol. Rev.*, 1962, **69**, 220–232.

MOORE, O. K., & ANDERSON, S. B. Modern logic and tasks for experiments on problem solving behavior. *J. Psychol.*, 1954, **38**, 151–160.

RAY, W. S. Complex tasks for use in human problem solving research. *Psychol. Bull.*, 1955, **52**, 134–149.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. USN Technical Report No. 11, 1954, University of Minnesota, Contract ONR N8 onr-66216, United States Navy, Office of Naval Research.

STAATS, A. W. Verbal and instrumental response hierarchies and their relationship to problem-solving. *Amer. J. Psychol.*, 1957, **70**, 442–446.

STORMS, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55**, 390–395.

(Received March 7, 1963)

Vol. 78, No. 4 Whole No. 581, 1964

Psychological Monographs: General and Applied

AN INFORMATION-PROCESSING MODEL OF ONE KIND OF PROBLEM SOLVING[1]

EDWARD S. JOHNSON[2]

University of North Carolina

Solutions to a set of concept-formation problems were secured individually from 11 Ss using a thinking-aloud procedure. The task required Ss to propose a sequence of hypotheses concerning properties which differentiate 1 set of patterns from another set. Verbal transcripts were analyzed in detail for consistencies in the ways in which Ss produce hypotheses. A method of low-level scanning was identified which forms the basis for generating simple hypotheses. More complex (disjunctive and conjunctive) hypotheses were found to be compounded from simple consistencies found in the pattern sets. Points of individual difference relating to the complexity of hypotheses formed, degree of validity checking, persistence, etc., were identified. An information-processing model was constructed which was judged to be successful in the simulation of problem-solving behavior of several Ss on a variety of problems.

THE last few years have witnessed an upsurge of interest in the study of cognitive processes. Evidence of this fact is found in survey articles of the last decade (Gagné, 1959; Johnson, 1950, 1955; Scheerer, 1954; Taylor & McNemar, 1955; Vinacke, 1952). Examination of this work shows that the dominant point of view remains essentially behavioristic. Traditionally the subject (*S*) has been viewed as a kind of black box and the experimenter's (*E*'s) task has been one of describing the relationship holding between what goes into and what comes out of the black box, i.e., between objectively measurable variables. Because organisms are inherently complex and because there is always a multiplicity of irrelevant inputs and outputs, investigators have usually limited themselves to the study of small numbers of variables at any one time. Consequently, models constructed from this point of view are characterized by small numbers of parameters and variables combined into rather simple functions. Furthermore, proponents of this approach would object to making use of *S*'s subjective reports because it is held that there is no necessary relation between this mode of behavior (input: question; output: answer) and the mode of primary concern (input: stimulus or problem; output: response or solution). Hull's (1920) classical study, for example, is often cited to show that *S* may respond as if entertaining a certain hypothesis while at the same time being unable to verbalize it. The objective approach is illustrated in a study by Bruner, Goodnow, and Austin (1957) who concerned themselves with strategies employed in certain types of problem solving. These strategies were identified by means of backward inference from objective performance measures.

Some investigators, prompted by interest in the process rather than the product of thought, have made use of subjective reports (for instance, Bartlett, 1958; Duncker, 1945; Wertheimer, 1945). This work did not, however, progress beyond the descriptive and speculative stage.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of North Carolina. The research was supported, in part, by Grant No. M-1849(c4) from the National Institute of Mental Health and by Grant No. AF-AFOSR-62-5 from the Air Force Office of Scientific Research.

The author wishes to acknowledge the support and interest of L. V. Jones under whose direction the research was conducted.

[2] Now a member of the Department of Psychology at Yale University. During the preparation of this paper for publication, the author held an appointment supported in part by a grant from the Ford Foundation and in part by Contract Nonr 609 (20) between Yale University and the Office of Naval Research. The grant and the contract provide support for a continuing program of research on thinking, program under the general direction of D. W. Taylor.

A new approach, which also makes use of subjective reports, regards *S* as a processor of information. The reports are used to shed light upon the nature of the information-handling processes employed by *S*. The *E* sets as his goal the construction of an artificial device or model which reproduces or simulates the relevant behavior exhibited by *S*.

It has been no accident that the development of the information-processing approach has followed upon the development of the digital computer as a modern research tool. Since, typically, an information-processing model is a set of highly complex logical and arithmetic functions, the implications of the model are best sought by operating it on a computer.

Although the information-processing approach has by no means completely proved itself, it already shows great promise. Miller, Galanter, and Pribram (1960), viewing man much as if he were a self-programed automaton, have investigated the nature of the program and the control system. Newell, Shaw, and Simon (1957, 1958, 1959, 1960) have investigated in a series of monographs the heuristics involved in human problem solving and have constructed a computer program for the solution of any kind of problem which can be subjected to a means-ends analysis. Feldman (1961) has simulated rather closely the behavior of individual *S*s in a binary-choice experiment. Feigenbaum (1961) has constructed an information-processing model which exhibits many of the phenomena of rote serial learning. Hunt (1962) in his recent book on concept learning leans heavily upon computer-oriented models.

The current study was intended as a further exploration into the feasibility of information-processing models of thinking and problem solving. A problem-solving situation was delineated, data were collected in the form of verbatim transcripts of problem-solving sessions, and a model was constructed on the basis of information contained in the data.

EXPERIMENT

The purpose of the experiment was to provide the information essential to the construction of an information-processing model. To construct such a model requires detailed information concerning the processes which occur. Several features of the experiment were dictated by the need to obtain this information.

First of all, *S* was directed to "think aloud" while trying to solve the problem. Because of the deficiencies of memory the record of *S*'s recollections after each problem had been solved would have been inadequate. The use of the thinking aloud procedure allows not so much *S*'s introspection on his own thought processes as it does a record of what he is thinking. This record, of course, is not complete. There is probably a level of detail beyond which *S* will not bother to report if only for the reason that the reporting of the behavior takes much longer than its performance. And, at the level *S* chooses to talk about, it should not be presumed that every step is reported. *S*'s verbalizations, then, will constitute the data of the experiment and will be taken as clues to or indicators of underlying information-handling processes.

Secondly, the type of task chosen to be employed was one which was segmented into parts, each of which gave *S* more information and which also required *S* to give some indication of the current state of his formulations. This feedback to *E*, in the form of a trial hypothesis, was useful in order to promote the thinking aloud procedure and also to insure that a close contact was maintained with *S*'s line of thinking.

Finally, as will become apparent, the procedure was such that once *S* had received a block of information, he had easy access to it thereafter. This relieved *S* of burdensome demands on memory so that more of his time would be spent on problem-solving processes of greater interest to the present study.

The following problem-solving task fulfills the above requirements. Consider a sequence of events or trials of which the following is a member:

● ● ○ ● ○

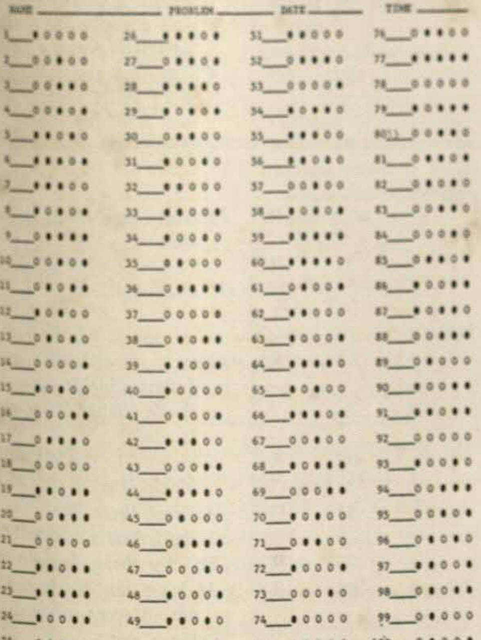
There are five channels each of which, on any given trial, is either black or white. Since there are five two-valued variables in a pattern (or, to use the terminology suggested by

Hovland, 1952, five dimensions, each with two possible values), there are 2^5 or 32 different patterns. To create a problem, the patterns were assigned to one of two disjoint subsets according to some membership rule. The patterns and their classifications were then presented to *S* sequentially (randomly with replacement). The *S*'s task was to discover the classification principle so that he could correctly predict the classification of the as yet unpresented patterns.

A problem sheet was printed consisting of a standardized sequence of 100 patterns. The sequence was random with the constraint that each of the 32 patterns must appear either three or four times. A further constraint that each pattern appear at least once in each half of the sequence was accidentally violated. (The pattern at Trial 68 appears there for the first time.) The problem sheet is reproduced in Figure 1.

Experimental Procedure

Each *S* was told at the time of recruitment that he could expect to spend two sessions of about 2 hours each in the experiment. At the beginning of *S*'s first session the following points of instruction were paraphrased by *E*.



NAME ____ PROBLEM ____ DATE ____ TIME ____

1____ ●○○○○
2____ ○○●○○
3____ ○○●●○
4____ ○○●○●
5____ ●●○●○
6____ ●●●○●
7____ ●●●○○
8____ ●○●○●
9____ ○●●●●
10____ ○○●○●
11____ ○●○●●
12____ ●○●○○
13____ ○●○●○
14____ ○○○○●
15____ ●○●○○
16____ ○○○●●
17____ ○●●●○
18____ ○○○○○
19____ ●●○●●
20____ ○○●●●
21____ ○○●○○
22____ ●●○●●
23____ ●●●●●
24____ ●○○●●
25____ ○●●●○
26____ ●●●○●
27____ ○●●○●
28____ ●●●●○
29____ ●○●○●
30____ ○●●○○
31____ ●○○●○
32____ ●●○○○
33____ ●●○○●
34____ ●○○●○
35____ ●●○○○
36____ ○●●●●
37____ ○○○○●
38____ ○●○●●
39____ ●●○○●
40____ ●○○○○
41____ ○●○○●
42____ ●●●○○
43____ ○○○●●
44____ ●●●●○
45____ ○●○○○
46____ ○●●●●
47____ ○○○●○
48____ ●○○○●
49____ ●●○●○
50____ ●○●●○
51____ ●●○○○
52____ ○●●●○
53____ ○○○○●
54____ ●○●●○
55____ ●●●○○
56____ ●●○●○
57____ ○○●○○
58____ ●○●○●
59____ ●●●●●
60____ ●●●●○
61____ ○●○○●
62____ ●●○○○
63____ ●○○○●
64____ ●●●●○
65____ ●○●○○
66____ ●●●○●
67____ ○○●○○
68____ ●○●●●
69____ ○○○●●
70____ ●○●○○
71____ ○●●○○
72____ ●○○○●
73____ ○○○●○
74____ ●○○○○
75____ ○●●●●
76____ ○●●○○
77____ ●●●●●
78____ ○○○○○
79____ ●○●●●
80____ ○○●●○
81____ ○●●○●
82____ ○●○●○
83____ ○○●●○
84____ ○○○●○
85____ ○●●○●
86____ ●○○●●
87____ ●○●●○
88____ ○○●●●
89____ ○●○○○
90____ ●○○●●
91____ ●●○●●
92____ ○○○○○
93____ ●○○●○
94____ ○○●●●
95____ ○○●○●
96____ ○●○●○
97____ ●●○○●
98____ ○●○●●
99____ ○●○○○
100____ ○●○○●

FIG. 1. The problem sheet.

> Here is a problem sheet. Note that there are numbered patterns of dots—100 in all. Each pattern contains 5 dots any 1 of which may be either shaded or blank. A problem consists of this. I have in advance decided upon a principle which will separate the patterns of dots into two groups—a red group and a green group. Your task is to discover this principle. You do this in the following way. I will tell you outright what the first pattern is, whether it is a green or a red. Mark through this pattern with the appropriate color [*E* demonstrates]. Then you predict what you think the next pattern is—red or green. After you have told me your prediction I will tell you the correct answer and you mark through with the appropriate color again. This procedure is repeated for each pattern in turn. Each time you will gain a bit more information and sooner or later you should discover the correct principle and be able to apply it successfully.
>
> Each time you make a prediction I want you to tell me what hypothesis you are working on. Between predictions I want you to think out loud as much as possible so that I can find out what ideas you are playing with, what hypotheses you reject, etc. Your whole object is to discover the principle so that you can use it to predict the color of all the remaining patterns. You will then receive a payoff according to the trial on which you first demonstrate that you are able to use the principle correctly. For instance, if this occurs on Trial 50 your payoff would be 25 cents. If it occurs on Trial 10 your payoff would be 45 cents.
>
> Here are 32 metal tags on which the various patterns are represented. You may use these in any way you wish or you may choose to ignore them. You may also use paper and pencil.

When it was clear that the procedure was understood, problems were administered for approximately 2 hours. A 5-minute break was allowed half way through each session. All sessions were recorded in their entirety, and the verbatim transcripts thus obtained constituted the data of the experiment. A representative transcript is presented in the Appendix.

Apparatus

The items of apparatus used in the experiment were: (*a*) the problem sheet containing 100 numbered patterns; (*b*) 32 small metal tags $\frac{5}{16} \times 1\frac{3}{4}$ inches, on each of which a different pattern was represented; (*c*) two felt-tipped markers with self-contained ink supply—one red and one green (the ink was diluted so that a marked-through pattern was still plainly visible); (*d*) pencil and note pad for *S;* (*e*) problem-scoring keys for *E;* (*f*) money (coins) used for rewarding *Ss*; and (*g*) tape recorder. All sessions were conducted in small experimental rooms devoid of major distractions.

Subjects

Ss were 11 male and female students enrolled at the University of North Carolina. Since *Ss* will occasionally be referred to individually, Table 1 gives the identifying initials and scholastic background of each.

Problems

Twenty-six different problems, listed in Table 2, were used in the experiment. A principle defines the members of one set of patterns. Thus all patterns not identified by the principle automatically belong to the other set. (Since *Ss* used red and green markers in the experiment, the principles are conceived as dividing the patterns into color sets.)

The problems may be classified into four groups. *Molecular* problems are based upon shadings of particular channels. *Numerical* problems concern the number of black dots in a pattern. *Molar* problems concern properties of patterns which are not confined to particular channels. *Sequential* problems are those in which the color classification of a pattern depends at least in part upon some characteristic of a preceding pattern. (Thus, sequential problems do not necessarily partition the patterns into disjoint sets.) It should be noted that this classification does not assign a problem to one and only one class. It is frequently possible, for instance, to express a molar problem as a molecular problem. This is easily seen with the molar concept of symmetry. Expressed in molecular terms the principle classifies together all patterns whose first and fifth channels are similarly shaded and whose second and fourth channels are similarly shaded. This dual nature of some problems is mirrored in different types of solutions attained by different *Ss* to the same problem.

Since one of the primary goals of the experiment was to insure richness of content in the data, the order of presentation of problems was decided ad libitum during the course of the problem-solving session. In this manner *S* could be presented problems which matched his growing competence or compensated for motivational or fatigue effects.

TABLE 1

IDENTIFICATION OF *Ss*

Initial	Identification
SB	Male undergraduate psychology major
FB	Male graduate student in computer-oriented mathematics
BC	Male undergraduate psychology major
KC	Male undergraduate psychology major
BG	Male prelaw undergraduate
SH	Male undergraduate psychology major
HL	Male undergraduate psychology major
TL	Male undergraduate psychology major
KR	Female psychology graduate student
AS	Female undergraduate psychology major
DS	Male psychology graduate student

TABLE 2

PROBLEMS USED IN THE EXPERIMENT

Problem type	Identification number	Color set identified by principle	Defining principle
Molecular			
	1	R	Channel 3 black
	3	G	Channels 1 and 5 both white
	4	G	Channels 2 and 5 shaded alike
	6	R	Channel 3 black and either Channel 2 or Channel 4 black
	8	G	Channels 2 and 5 both white
	9	G	Channels 1 and 5 shaded alike
	17	R	Channels 2 and 4 both white
	29	R	Channel 3 black and Channels 1 and 5 alike or Channel 3 white and Channels 1 and 5 different
	30	G	Channel 1 shaded differently from either Channels 2 or 4
	31	R	Channel 1 black, Channels 2 and 4 alike, and Channels 3 and 5 alike
	32	R	Channels 1 and 4 shaded alike and Channels 2 and 5 shaded alike
	33	R	Channels 1 and 4 alike
Numerical			
	2	R	Exactly two black channels
	22	R	An even number of black channels
Molar			
	5	R	Exactly two shade runs (alternatively: one shade change from end to end)
	12	R	Weighted to the left with black dots
	15	G	At least three like-shaded dots adjacent
	20	R	Exactly three shade runs
	24	R	At least one black dot not adjacent to another black dot
	25	G	Symmetrically shaded

TABLE 2—Continued

Problem type	Identification number	Color set identified by principle	Defining principle
Sequential			
	7	G	Channel 1 of the preceding pattern is black
	11	G	The first pattern with Channel 1 black; thereafter, Channel 1 black and Channel 5 shaded differently from Channel 5 of the preceding green pattern
	19	R	The same number of black dots as the preceding pattern.
	28	R	The first pattern with one black dot, the next pattern with two, the next with three, the next with four; thereafter, the cycle repeats
	34	G	The preceding pattern has one black dot
	35	G	The first pattern; thereafter, a pattern is green if it is shaded differently in Channel 1 from the preceding green
	36	—	The first pattern is green; thereafter, if a pattern has four black dots it belongs to the opposite color set from the preceding pattern; otherwise, it belongs to the same color set

SOME PRELIMINARY FINDINGS

The primary data obtained in the experiment were the verbatim protocols. However, before a discussion of these data is presented it is desirable to examine some supplementary data in order to provide a framework within which that discussion may be placed.

Table 3 lists the problems in the order in which they were presented to each *S*. The column headed Trial gives the trial number on which a correct hypothesis was first verbalized or, in the case of unsolved problems (indicated by an asterisk), the trial on which either *S* or *E* called a halt. The column

TABLE 3

PERFORMANCE MEASURES ON PROBLEMS

Subject	Problem	Trial	Time
SB	2(1)	24	24
	15	16	16
	25	22	35
	12*	38	78
	1	8	6
	2(2)	23	13
	3	46	53
	32*	31	35
FB	1	3	2
	8	9	11
	20	50	43
	12	31	—[a]
	32	28	15
	22	5	4
	28*	18	90
KC	1	6	12
	2	6	5
	8	6	7
	9	2	4
	3	8	8
	20	48	60
	17	6	5
	32	28	12
	12	70	55
	24	50	36
BC	1	7	4
	2	42	20
	3	13	8
	12	34	15
	15	24	20
	22	15	5
	5	30	15
	20*	30	60
BG	1	8	8
	2	6	5
	3	21	58
	15	18	23
SH	1	14	10
	2	11	7
	3	9	12
	4*	23	55
	6	9	10
	7*	17	—[a]
	33	9	10
	30	9	7
	29	20	9
	31	24	26
	32	42	24
	28*	16	40
HL[b]	25	6	9
	12	9	9

TABLE 3—Continued

Subject	Problem	Trial	Time
TL	1	7	8
	2	18	32
	3	7	10
	4	8	23
	22	18	8
	24	19	24
	15	21	12
	5	19	20
	19*	25	104
KR	1	15	14
	2	13	6
	3	11	9
	4*	28	70
	5(1)	24	20
	12	31	20
	5(2)	20	23
	22	41	48
	24	50	30
AS	1*	50	45
	2	13	8
	3	9	8
	4(1)*	50	50
	12	14	5
	15*	50	38
	22	7	4
	4(2)*	23	—[a]
DS	28*	100	62
	19*	72	69
	34	19	10
	35	34	31
	11	36	44
	36	37	21
	15*	38	42

Note.—Problems are listed in the order in which they were presented to *S*. Unsolved problems are indicated with an asterisk. The numbers 1 and 2 in parentheses are used to denote the first and the second presentation of the same problem to the same *S*.

[a] Time not available due to faulty equipment.

[b] Performance on several problems lost due to faulty equipment.

headed Time gives the total time, in minutes, spent on the problem. This includes not only the time required for solution, but time spent (*a*) in going several trials beyond the trial given in the preceding column in order to insure that *S* could successfully apply his rule, (*b*) in a brief postproblem discussion, and (*c*) in awarding the payoff.

A rough comparison of the difficulties of problems can be derived from a tally of the number of times a solution was not reached. Sequential problems were missed 6 out of 10 times. (The 4 successes were accounted for by one *S*.) As a class, numerical problems (14 solved out of 14 presented) seemed easier than either molecular problems (29 solved out of 35 presented with AS accounting for half the failures) or molar problems (21 solved out of 24 presented). But it should be remembered that the distinction between problem classes is sometimes vague.

Various other performance measures on the problems, broken down according to problem type, are given in Table 4. The amount of time spent on a single problem ranged from 4 to 104 minutes with an average of 18 minutes for solved problems and 54 minutes for failed problems. The number of trials to solution ranged from 2 to 72. The average point of solution was Trial 20.

HYPOTHESES

As was mentioned above the primary data of the study were the verbatim protocols. It is from these protocols that an understanding must be gained of the problem-solving processes used by *S*s. After intensive examination of the protocols (see Johnson, 1961, for details of this examination), these processes were found to fall into two main categories: a set of specific processes for constructing individual hypotheses and a set of broader strategical processes which serve to direct the search for a solution.

As would be expected, the working hypotheses which *S*s were required to verbalize during the course of a solution were based upon those attributes of patterns which they noticed. By and large, all *S*s noticed the same attributes; and their classification, as given in Table 5, easily follows that of problems except that within each class two or more subclasses can be identified. Hypotheses, then, consist of unitary attribute-value pairs used alone or combined by using the logical operators "and" and "or." Briefly, *S*s noticed: (*a*) the shading of particular channels, (*b*) the similarity of shading of pairs of channels, (*c*) the number of dots of similar shade, (*d*) the number of dots of the majority shade, (*e*) symmetry, (*f*) weighting, (*g*) run lengths of like shaded dots, and (*h*) the adjacency of like shaded dots. No sequential attributes are listed be-

TABLE 4

SUMMARY OF PERFORMANCE MEASURES ON PROBLEMS

Problem type	Number attempted	Number solved		Mean number of trials per problem	Mean time per problem (in minutes)	Mean time per trial (in minutes)
Molecular	35	29	solved	14	10	.71
			failed	34	51	1.46
Numerical	14	14	solved	17	14	.82
			failed	—	—	—
Molar	25	21	solved	29	24	.83
			failed	39	54	1.38
Sequential	10	4	solved	32	26	.84
			failed	41	64	1.56
Overall	84	68	solved	20	16	.80
			failed	38	57	1.50

cause of the failure of all but one *S* to formulate valid sequential hypotheses.

In addition to the attributes listed in Table 5 *Ss* used other information to construct hypotheses. "Color" hypotheses are of a sequential nature but are concerned not with any property of the patterns but with the sequence of color classifications. The *S* might have noticed, for instance, that every third pattern had been classified green and predicted accordingly. Color hypotheses were used with only moderate frequency and usually only during the first 5 or 10 trials of a problem.

"Actuarial" hypotheses make use of redundancies in the sequence of patterns. The *S*, in trying to predict the color class of a pattern, might examine the already classified patterns for a prior occurrence of the same pattern, a prior occurrence of its complement (a black or white reversal), or a prior occurrence of its converse (a left to right reversal). This information was sometimes relevant, sometimes not, but in either case *S* was attempting to increase the likelihood of a correct prediction at the expense of coming no closer to a true solution.

Notation

Since many hypotheses will be mentioned in later sections, it is desirable at this point to provide a shorthand method for writing them. Table 5 lists mnemonic symbols for attributes (digits or capital letters) and values (digits or small letters). Attribute value pairs are written in the order attribute-value and are separated by a colon. For example, 23:s means that Channels 2 and 3 are similarly shaded. Hypotheses are symbolized by the letter R or G followed by one or more attribute-value pairs. Several attribute value pairs are combined using the logical operators "and" (·) and "or" (v). Brackets and parentheses indicate the scope of the operators. Thus the hypothesis "the reds either have three black dots or they have two black dots, one of which must be at the left end" is written R:: NB:3 v [(NB:2)·(l:b)].

STRATEGY

Low-Level Scanning

While attempting to solve a problem, *S* scans over the members of the two color sets in the effort to find a property (an attribute-value pair) shared by all the members of one color set and by none of the members of the other. In an attempt to discover the scheme used by *Ss* in selecting particular attribute classes for scanning, sequential lists were constructed of all attribute classes and subclasses mentioned by *Ss*. The result was that no well-defined scheme was discernible. Rather, *Ss* seemed to be following a procedure of tending to investigate again those attribute classes which have previously succeeded in providing working hypotheses and, conversely, to neglect those which have failed. This is born

TABLE 5

NOTATIONAL SYSTEM FOR THE ATTRIBUTES AND VALUES NOTICED BY *S*s

Attribute class	Attribute subclass	Attribute notation	Values noticed	Value notation
Molecular	One channel (shade of Channel i, $i = 1, \ldots, 5$)	i	Black, white	b, w
	Two channels (shade of channels i and j, $i \neq j = 1, \ldots, 5$)	ij	Same, different, at least one black	s, d, b
Numerical	Number (number of dots of Shade i, $i = $ B, W.)	Ni	None, one, two, three, four, five	0, 1, 2, 3, 4, 5
	Plurality (number of dots of majority shade)	P	Three, four, five	3, 4, 5
Molar	Balance	B	Symmetric, weighted	s, w
	Weighting	W	Right, left	r, l
	Run length (length of longest run of Shade i, $i = $ B, W.)	Li[a]	One, two, three, four, five	1, 2, 3, 4, 5
	Adjacency (togetherness of dots of Shade i, $i = $ B, W.)	Ai[b]	Together, separated (not all together)	t, s

[a] If i($=$ B or W) is omitted then the longest run regardless of shade is indicated.
[b] If i($=$ B or W) is omitted then adjacency of both shades simultaneously is indicated.

out by calculation of transitional probabilities. The probability of *S*'s scanning within the context of a particular attribute class, given that he had just conducted an unproductive scanning elsewhere within it, was .13 while the probability of returning after scanning in another class rose to .33.

Since no further details of search were discovered, low-level scanning will be described as an essentially random procedure but with constraints which tend to give direction to the search.

Complex Strategies

More complex strategies were occasionally used by *S*s. These strategies involve the combination of separate bits of information, found during low-level scanning, into conjunctive, disjunctive, or conjunctive-disjunctive hypotheses. Two conditions were usually necessary before *S* would attempt to form a complex hypothesis. First, low-level scanning must have failed to solve the problem. (Usually *S* had proceeded beyond Trial 10 and no solution was in sight.) Second, *S* must have discovered, or think he had discovered, a property in one of the color sets which was either a necessary condition or a sufficient condition.

If a necessary condition was detected (i.e., all items of one color set have a property, p, and some items of the other color set also have p), one course of action was: scan the same color set for another necessary condition, x, to be combined conjunctively with p to form the hypothesis: H:: $p \cdot x$. This was the single most common complex strategy used by *S*s. Usually no comment was made concerning the fact that, having found one necessary condition, another one was needed to make the hypothesis sufficient as well. Instead the necessary condition was noted, followed quickly by the completed hypothesis. Occasionally the details of the search were more explicitly verbalized. For example, SH (Problem 31, Trial 24) said:

> In all the red ones the third and fifth dots are the same color. All right, but that's true in some of the green ones, too, so there must be something else . . . etc. . . . all the red ones have an odd number of dots too . . . an odd number of colored dots . . . some of the green ones do . . . all right . . . let's see what could I add to this . . . an odd number of colored dots and the second and third dots are the same color . . . let's see, some of these greens—no, there's a green one that's the same way . . . An odd number of colored dots, and the third and fifth one are the same color . . . .
> I'm getting more complicated on this mess. All right, I'm trying to add up all the things that look alike in the red ones and see if I can come out with something, but I'm getting wor—all right, they got an odd number of colored dots, the third and fifth one are the same color and the first one is black. I'd better go on with that before I forget it.

The other possible action, having detected a necessary condition, was: scan the items in the other color set which also have p for another property, x, not possessed by any items in the first set. The hypothesis is then of the form: H::p unless x. (Although "unless" is logically equivalent to "and not," *S*s never used the latter term.)

Sufficient conditions (i.*e*., all the items which have a given property, p, belong to one set, but there are other items without p in the same set) were somewhat less commonly detected. This may be the consequence of low-level scanning in which *S* searched for a necessary and sufficient condition. That is, *S* tried to find a property common to all the elements of a color set. If *S* then scanned the other color set and found that the property occurred there also, he had found a necessary condition. The discovery of a sufficient condition, on the other hand, could arise in two ways. First of all, *S* might have discovered that all items of one color set lacked a certain property. This could have led to the discovery that the property was a sufficient condition for membership in the other color set. Alternatively, *S* might have concerned himself, not with finding a property common to all items of a set, but rather with the distribution of items having a particular property (e.g., "Where are the symmetric patterns now?"). This could have led to the discovery of a property which occurred in just one of the color sets.

In any event, if a sufficient condition was detected the following action might be taken: rescan the same color set for another property, x, to be combined disjunctively with the original property to form the hypothesis: H::$p \vee x$. For example, FB (Problem 28, Trial 7) said:

> . . . The characteristic of some of these is that they all have at least two dots adjacent, two blue dots adjacent. Not this one . . . . color changes won't work either . . . count of the number of changes of color in each one . . . . . . Well, I'll go along with the rule I started to state. If it contains two blue dots in a row then it's red, or if it contains four white dots in a row which—there're only two patterns which do this—then it's red. Otherwise green.

Another strong clue which was often noticed in the data is an approximate necessary and sufficient condition. It holds when almost all items with a property, p, belong to one color set. However, some items in the set lack p and some items in the other color set have p. (This will be referred to hereafter as a majority condition.) It would seem reasonable for a majority condition to be the basis for the types of hypotheses, H:: $(p \vee x)$ unless y or H:: (p unless y) $\cdot$ x, where x and y are properties which serve to account for the exceptions to the rule H:: p. Although the data failed to show a single case of this particular strategy, majority conditions still served to stimulate the formation of hypotheses. Usually the majority condition was handled as if it were either a necessary condition or a sufficient condition (and one of the above mentioned strategies is employed) or a necessary and sufficient condition. It is not clear whether an accidental misperception was involved or the data were intentionally simplified by ignoring inconsistencies.

Once a working hypothesis had been generated, it would most likely meet with a contradiction sooner or later. The *S* was then faced with the alternatives of either completely rejecting the hypothesis or attempting to salvage and repair it. Here again, complex strategies can come into play, for a contradiction turns a necessary and sufficient condition into just one or the other. The following cases arose.

If a simple hypothesis is contradicted and the exception is in the same color set, then the original hypothesis is now just a sufficient condition. It can be repaired by adjoining another sufficient condition which distinguishes the exception from all the items in the other color set. If the exception is in the other color set, then the original hypothesis is now just a necessary condition. It can be repaired by either adjoining another necessary condition conjunctively or by using an "unless" clause which distinguishes the exception from the items in the original color set.

If a complex hypothesis is contradicted, the above strategies may be employed to adjoin new conditions which account for the exceptions. Another method is to delete the contradicted element (thus creating several more exceptions to the reduced rule) and account for all the exceptions with a new element.

The above discussion should not leave the impression that all *Ss* commonly employed complex strategies. BC, BG, HL, and AS jointly accounted for just five instances. At the other extreme, however, KR made extensive use of complex strategies on all problems after the first one, while FB, SH, TL, and KC averaged about a dozen occurrences apiece. Furthermore, those *Ss* who were more likely to construct complex hypotheses were also more likely to attempt the repair of a refuted hypothesis. Once a hypothesis had been thus repaired, it became very hard for *S* to ever reject it. Instead he kept on repairing it as long as he could find ways in which to do so.

MALADAPTIVE BEHAVIOR

Most problems were not solved in the easiest and quickest manner possible. Behavior which tended to detract from an otherwise optimal performance was exhibited by all *Ss* on most problems.

Incomplete Validation

Whenever *S* mentioned a possible hypothesis he usually made some attempt at verification before offering it as his next working hypothesis. However, some 28% of all working hypotheses were invalid (with a range for *Ss* from 5 to 56%). In most cases the number of contradictions overlooked was very small, usually just one or two. The effect which an invalid hypothesis had upon *S*'s progress toward a solution was generally slight since it seldom predicts successfully for very many trials. The effect, however, could be more serious if the invalid hypothesis was close to a correct solution.

Perseveration

There is a set of behaviors whose common theme is a tendency to persist too long in a given line of attack. It is a reluctance or inability to change in the face of continued frustration and failure. There are several classes of this type of behavior.

1. Attention is focused on the "wrong" color set. It is possible to solve any nonsequential problem in two ways—characterize the red set or characterize the green set. In some problems, especially molar problems, there is often an easier set upon which to focus attention. This set may be much "simpler" perceptually. Much of the trouble experienced on Problem 12 (R:: W:1), for instance, was due to the fact that attention was directed solely toward the green set which included all the symmetric items as well as the right-weighted items.

2. A hypothesis is repaired repeatedly instead of being rejected. This behavior was the dominant characteristic of KR on her later problems. Other *Ss* showed this to a much lesser degree. Logically, this procedure cannot fail if pursued long enough. The drawback is that simpler solutions will go unnoticed.

3. Old methods are used in new situations. The best examples of this behavior occurred when *Ss* were faced with sequential problems. Without exception, *Ss* recognized the sequential aspect of the problem. However, the universal response was to continue searching for a nonsequential solution. The only exception was DS in his later problems.

Failure to Reach a Solution

Of the 84 problems attempted by *Ss*, 16 were not solved. Three *Ss* jointly accounted for 10 of these failures. Four *Ss* had one failure apiece and three *Ss* had no failures. Six of the failed problems were sequential, 6 were molecular (including 4 failures on problem 4), and 4 were molar. No numerical problems were failed.

In many instances of failure the attribute required for solution had not been mentioned by *S* on the failed problem or on any previous problem. Such a gap in *S*'s repertoire of attributes made it impossible to solve certain kinds of problems. This critical-factor explanation applies to all four of the problems missed by AS. In no case had she ever before used the type of hypotheses necessary to solve the missed problems. It also applies to the failures of KR and SH on Problem 4 and SB on Problem 32.

This explanation is not applicable to all problems. Problem 12 (R:: W:1), for instance, can be solved without using the concepts of weighting and symmetry. The process is, however, quite difficult and time consuming. The same applies to molar problems involving length of run, number of runs, symmetry, etc. It would appear that the critical factor aspect applies only to molecular problems. If a necessary molar concept was unavailable to an *S*, then his success hinged on whether or not he was able to use a complex strategy to put together a complex molecular hypothesis. (Note that the single-channel subclass of attributes can be used alone to solve any nonsequential problem since this set of attributes is sufficient to specify the patterns uniquely).

It is now possible to give an overall picture of the typical solution. All *S*s employed a technique of early hypothesis selection which tended to insure that an easy problem would not be misinterpreted as a hard problem because of an unnoticed but simple solution. This technique called for the rapid use of a series of simple hypotheses early in the problem, each one being used until it was refuted. At some point, usually between Trials 5 and 10, the problem had either been solved by a simple hypothesis or the search slowed down greatly as *S* began to investigate more complicated relations. It is doubtful if *S*s were consciously guarding against being fooled by simple solutions. It is more likely that they were waiting for the accumulation of a body of data before launching serious searching. In the meantime they fulfilled the requirement of justifying each prediction.

It is interesting to notice the results when the transition from simpler to more complex searching failed to occur or occurred prematurely. AS could apparently perform only simple low-level scanning. (She produced only one complex hypothesis and that was based on a pet attribute which she often examined.) As a result, any problem not solved by Trial 14 was not solved at all. Beyond this point a series of actuarial predictions and guesses was used. Thus AS acted as if she perceived two basic types of problems, one kind having easy solutions and the other kind solvable only by using repeated patterns, reversed patterns, etc.

On the other hand, KR, during her second session, characteristically began complex scanning prematurely. For instance, her hypothesis on Trial 6 of Problem 24 was: R:: [(23:d)·(34:d)] v [P:4]. This premature complex scanning plus her tendency to repair refuted hypotheses rather than reject them led her to miss the simple solution to Problem 22 (R:: NB:even) and finally arrive after 39 trials and 48 minutes at the logically equivalent solution: R:: [NB:0] v [(NB:4)·(P:4)] v [(NW:3)·(P:3)]. No other *S* required over 8 minutes or 18 trials to solve this problem.

Once *S* made the transition to a more complex search, he tended to exhibit characteristic behavior which set him apart from other *S*s. The *S*s varied, for example, in the degree to which they followed up the discovery of a necessary condition or a sufficient condition with an attempt to form a complex hypothesis, in the number of invalid hypotheses used, and in the frequency with which they employ guessing.

In the light of this variation among *S*s, it would be useful to know something of the variation within a single *S*. Three *S*s received the same problem twice, each time apparently failing to recognize the second presentation as a repeat. AS failed Problem 4 twice, both times in a manner characteristic of all her failures. KR solved Problem 5 twice using the same basic approach. SB solved Problem 2 twice and in identical ways even down to making the same type of comment at corresponding points in both solutions. The implication of these three repetitions seems to be that something which might be termed a basic approach or style of solution remained fairly constant across solutions by the same *S* but that the particulars of a pair of solutions to the same problem were apt to differ. This observation will have relevance when performance of the model is considered.

MODEL

Before the construction of a model for the simulation of problem-solving behavior can begin, several basic questions of goal and intent should be answered. What is the model to accomplish? At what level of behavior is the simulation to be carried out? What behavior can safely be ignored as being irrelevant to the study? The answers to these questions are somewhat arbitrary and tend to reflect not only the personal biases of the model builder but also the extent of knowledge which he has of the behavior to be simulated.

There are perhaps three types of information processing models which can be constructed. By far the easiest type would be an "optimal" model which would not intentionally simulate, but instead would solve the various types of problems in as simple and efficient manner as possible. Such a model would use a set of strategies devised by the model builder. The data would serve merely to suggest possible strategies which might be employed. This type of model is a simulation model only in that it should perform in a manner similar to an intelligent and highly experienced *S*. The Logic Theorist of Newell, Simon, and Shaw would fall into this category.

An optimal model should not be confused with an algorithm. An algorithm will, by definition, eventually come to a solution, but will not intentionally simulate human behavior (see, for instance, Wang's, 1960, algorithm for solving logic theorems). An optimal simulation model makes use of successful methods employed by many different *S*s. It provides no guarantee of solution; but if a solution is reached, it may often be achieved with less time and effort than the corresponding algorithmic solution.

Since the optimal model is a problem solver rather than a behavior simulator, its success is gauged purely according to its ability to solve problems. This criterion is decidedly different from the criteria used in judging the performance of the other types of models.

A second type of model would simulate a particular *S*. Such a model would be constructed using operations and strategies which this *S* actually employed as well as the various maladaptive features such as incomplete validation of hypotheses which characterize his problem-solving behavior. The binary-choice models constructed by Feldman (1961) belong to this category. A test of the success of this "individual" type of model, following Turing's (1956) criterion, would probably be carried out as follows. The same problems which *S* solved would be solved by the model. Since, at this stage, the two types of solutions are in different forms, (a tape recording as opposed to a list of hypotheses in symbolic notation) the two kinds of transcripts would have to be reduced to some common form such as, say, simple English declarative sentences. The two sets of transcripts would then be intermixed and given to a number of judges whose task would be to sort them out into two stacks again, according to whether they judge them to be of human or model origin. The model's success is gauged by the extent to which this separation of transcripts results in a random arrangement.

The third type of model is the most difficult to construct. It would bring together several individual models each of which simulates a different *S*. Ideally this "universal" model is a single system rather than separate systems for separate *S*s. The points of individual difference are given by parameters whose values represent degrees of specific kinds of behaviors. Thus, any *S* could be characterized by a particular set of parametric values which, together with a set of problems, constitute the model's input. The output is a transcript simulating the given *S*. The model is evaluated in a manner similar to the evaluation of the individual model.

A second major question is this. Shall the model be probabilistic or deterministic? In order to appreciate the distinction, consider the task of the model builder. He may choose to regard *S* as if he were a problem-solving automaton operating on the basis of a large set of variables. *E*'s task is to identify a sufficient number of relevant variables so that he may construct another automaton (cf. model, theory, computer program) whose behavior will closely match certain aspects of *S*'s behavior. Some variables are not considered because they are usually psychologically irrelevant (e.g., state of health). Still others may be psychologically relevant but of little consequence.

When *E*'s model is operated, the model-environment system (in this case the environment is a sequence of patterns constituting a problem along with their classifications) proceeds from state to state with each state leading to its successor state with unit probability. If a certain sequence of states of the system, S_i, S_j, S_k, is observed, then *E* knows that if ever S_i recurs it must of necessity be followed by S_j and S_k. However, if *S*'s behavior is examined (and if a state is defined over the same subclass of variables as for the model), it will be found that S_i will not lead to S_j with unit probability. The apparent randomness observed in *S* is due to the effect of variables which *E* chooses to ignore.

The *E* may now take one of two courses of action. He may choose to take the discrepancy between his model and *S* as a measure of the adequacy of the particular subset of variables he has selected. Alternatively, he may add a random component to the model with hopes that it will itself simulate the effect of the unincluded variables. Here too, discrepancy can be regarded as a measure of the adequacy of the selected set of variables.

The model discussed in the following sections is probabilistic. However, it will be pointed out how very minor changes can make the decisions involved completely deterministic. Although the model has not been operated deterministically, the author sees no reason why the output should differ materially from that of the model presented.

The last major question involves the extent of simulation. Just what is to be included within the domain of simulated behavior? There are several alternatives. The data of the experiment consist of verbatim transcripts. The ultimate in simulation—reproduction of the transcripts word for word—is out of reach and, in fact, includes many problems of perhaps greater complexity than the simulation of problem solving per se (e.g., the generation of English sentences and phrases). It is the process of problem solving which is of interest here, and this can be studied without concerning oneself with the particulars of the language in which it is encoded. Furthermore, an examination of the transcripts reveals that some *S*s are much more thorough in their reporting than are others. SB's data, for example, are filled with much dead space during which he was apparently engaging in data processing, most of which went unreported. On the other hand, TL talked constantly, and his data presumably represent a much more complete record of his problem-solving behavior. The data of primary interest, however, are consistently available from each *S*, that is, a justification for each prediction. The rest of the data were of use in gaining insight into *S*'s problem-solving behavior, but need not be reproduced by the model. Hence the data to be simulated consist of a sequence of hypotheses, one for each trial, starting with Trial *2* and continuing until either the problem was solved or *S* gave up. (An alternative would be to impose an output parameter on the model indicating the extent to which it reports on its own search behavior.)

One set of behaviors about which little information was gained involves sequential hypotheses and problems. Only one *S* yielded worthwhile data on the solution of sequential problems, and even these data are not too revealing. The other *S*s who attempted to solve sequential problems succeeded in showing merely that they cannot ordinarily be solved without prior experience. Consequently, no attempt will be made to simulate this type of behavior.

Basic Model

This section will outline a model which will perform what has been termed low-level scanning. On the basis of this scanning, the model will be able to generate *simple* hypotheses, that is, hypotheses based upon one or more values of a single attribute. In certain respects the model will be optimal (i.e., it will solve certain problems in a very efficient manner), and in other respects it will be individual (i.e., certain elements of the model will operate in the way in which *S*s are presumed to operate). The model will then serve as a basic framework to which will be added, in succeeding sections, provisions for conjunctive strategy, disjunctive strategy, and maladaptive behavior.

The solution of problems of the type employed in this study involves the scanning of individual patterns of a color set with respect to a particular attribute. The *S* looks for values which are common to all the patterns of the color set being scanned. Any such value found can become the basis for a working hypothesis providing a scanning of the other color set substantiates it. For instance, *S* may scan the red set with respect to number of black dots and find that all patterns have exactly two. If the same attribute-value pair is not now found to occur in the other color set, the current working hypothesis becomes R::NB:2.

The *S*s also frequently compound certain values of the same attribute. For instance, the values of the attribute "number of black dots" are the numbers 0–5. If the red set contains only patterns with values of 3, 4, or 5 (and these values do not occur in the other color set), then the hypothesis R::NB:2 ≥ 3 can be formed. The attributes and values noticed by *S*s were given in Table 5. The sets of values of a single attribute from which the model can generate simple hypotheses are listed in Table 6. The specification of these value sets enables the model, as described thus far, to generate nearly all of the simple hypotheses encountered in the experiment.

The formation of color hypotheses, i.e., those which are based on the sequence of pattern classifications rather than on the patterns themselves, requires a separate scanning device, one which is able to detect simple sequential and cyclic patterns. The input to this device is a string, composed of the elements R and G, which encodes the classifications of the thus far presented patterns in the order of presentation. This string is in turn encoded into a string of digits which gives the lengths of successive color runs. Cycles are then sought by a simple comparison technique which begins by searching the string for a repeated instance of the first element. The elements following the repeated instances are then compared with the second element. The process is continued until it is found that no pattern is present or until one is found which accounts for all the elements of the string (with the possible exception of the last element which might represent the length of an incomplete run; i.e., the next, as yet unknown, pattern may turn out to be in the same color set as the last pattern). Simple sequential patterns (arithmetic progressions) in the string of numbers can be found by forming differences between adjacent pairs of numbers and checking to see if all differences are equal.

TABLE 6

SETS OF VALUES OF A SINGLE ATTRIBUTE WHICH PRODUCE HYPOTHESES

Attribute class	Attribute subclass	Components of subclass	Values noticed	Hypothesis-producing subsets
Molecular	One channel	1, 2, 3, 4, 5	b, w	b, w
	Two channel	1–5, 1–2, 4–5, 2–3, 3–4, 1–3, 3–5, 2–4, 1–4, 2–5	s, d, b	s, d, b
Numerical	Number	Black dots only, White dots only	0, 1, 2, 3, 4, 5	0, 1, 2, 3, 4, 5 Odd numbers Even numbers Numbers ≧ 3 Numbers ≦ 2
	Plurality	The whole pattern	3, 4, 5	3, 4, 5
Molar	Balance	The whole pattern	s, w	s, w
	Weighting	The whole pattern	r, l	r, l
	Run length	Black dots only, White dots only, Both shades	1, 2, 3, 4, 5	1, 2, 3, 4, 5 Lengths ≧ 3 Lengths ≦ 2
	Adjacency	Black dots only, White dots only, Both shades	t, s	t, s

Note.—Whenever a particular scanning reveals that all the patterns of one color set have attribute values contained in one of the subsets listed in the right-hand column and that none of the items in the other color set have values in that set, then a hypothesis can be formed.

The component of the model which selects the scanning context remains to be specified. The results given in the section on low-level scanning suggest a priority listing for attribute classes with the class at the top of the list having the greatest probability of being selected for scanning, the probabilities decreasing monotonically to the bottom of the list. If a scanning yields a working hypothesis, the position of the attribute class involved is raised on the list; otherwise, it is lowered. A mathematically equivalent device (similar to functions used by, e.g., Bush & Mosteller, 1958, p. 8) assigns probability weights, w_i, to the attribute classes. If a scanning of attribute Class i produces a hypothesis, then compute a new weight w_i' for the class where

$$w_i' = \phi_1 w_i + (1 - \phi_1), \qquad [1]$$

and where w_i is the weight prior to scanning and ϕ_1 is a parameter in the range 0–1. All other weights are adjusted downward according to the rule

$$w_j' = \phi_1 w_j \text{ for all } j \neq i, \qquad [2]$$

in order to preserve the restriction that the weights sum to unity.

If a scanning of Class i fails to produce a hypothesis, then the weight w_i is reduced according to the rule

$$w_i' = \phi_1 w_i, \qquad [3]$$

and the weights of the other classes are increased proportionately according to the rule

$$w_j' = \frac{(1 - w_i')w_j}{1 - w_i} \text{ for all } j \neq i. \qquad [4]$$

The parameter ϕ_1 is, so far, the only parameter of the model. Its effect can readily be seen from the above equations. If ϕ_1 is near unity, the success or failure of a particular scanning will have little effect upon the set of weights. If it is near zero, the effect will be much more extreme. A low value of ϕ_1 would indicate an *S* who skips around readily from one type of attribute to another, while a high value would indicate an *S* who is more apt to explore a given attribute class thoroughly at the risk, perhaps, of neglecting other classes.

Within each attribute class there are two or more subclasses. The *S*s not only select attribute classes to investigate, but within the class select a subclass. For instance, the molecular class is not chosen for scanning but rather the subclass composed of pairs of channels. However, the two molecular subclasses do not function completely independently of one another; for if one subclass is successful in producing a hypothesis, the associated subclass shares in the success, i.e., the weight of the class as a whole is increased. (There is, however, some justification for doubting that the four molar subclasses hang together in the same way as do other sets of subclasses.) Thus each subclass also has a probability weight, and they are manipulated in the same way as class weights and follow the restriction that the weights within a class sum to unity.

The above discussion of the scanning device and its input selector are almost sufficient to specify a very simple problem solving model. Although the model is described as a computer program, it has not actually been programed,[3] but rather only "hand simulated" according to the set of rules. Therefore there are certain rather trivial human operations (for instance, applying a trial hypothesis to a color set to check the validity of the hypothesis) which would have to be specified in detail for a computer, but which are omitted from the present discussion. For the most part, these operations would present only routine programing difficulties. The model works according to the following set of rules.

1. An attribute class is chosen at random for scanning with each class having a probability of selection equal to its probability weight. A subclass of the chosen class is selected in the same manner. (The corresponding deterministic decision rule would be: always select that class and within it that subclass with the largest weight. The weights are changed as before with Equations 1–4.)

2. Having chosen a subclass, both color sets are scanned. If a subclass is composed of several components (e.g., the two-channel subclass is made up of 10 attributes—one for each possible pair), they are scanned in the prescribed order as given in Table 6.

---

[3] During the past year the author has been programing the model for the IBM 709 using IPL-V, a list processing computer language currently available for several types of computers. Preliminary results confirm the findings reported in this monograph.

3. Whenever the scanning of a subclass yields a working hypothesis, the probability weights of class and subclass are boosted by means of Equation 1, and the weights of the other classes and subclasses are lowered using Equation 2.

4. All hypotheses are checked for validity, and invalid hypotheses are immediately rejected. The first valid hypothesis found is the one used as the working hypothesis. It is retained until it either solves the problem or is refuted.

5. Whenever a subclass is scanned and no hypothesis is found, the probability weights of class and subclass are lowered by means of Equation 3. The weights of the other classes and subclasses are adjusted upwards with Equation 4.

6. If a specified number (given by the parameter α) of scannings fails to give rise to a valid hypothesis, search halts and an actuarial hypothesis is attempted. In order of priority actuarial predictions make use of repeated, converse, and complementary patterns. If no actuarial hypothesis is possible, then a random guess is made.

7. Upon refutation, the current working hypothesis is rejected, and scanning resumes in the same subclass from which it arose. If the subclass does not produce another hypothesis, it is treated as in Rule 5.

8. A "color" class is included in the set of attribute classes. Whenever it is selected for scanning, the model will attempt to form a color hypothesis according to the procedure outlined on page 14.

9. Before a compound hypothesis (i.e., a hypothesis involving several values of the same attrbute) can be formed, there must be at least $n + 1$ members in the scanned color set (where n is the number of unique values involved) and at least one member in the other color set. This rule prevents the overuse of compound hypotheses early in a problem.

The general plan of the model is given in flow-chart form as Figure 2. An example of the operation of the model is given in Table 7.

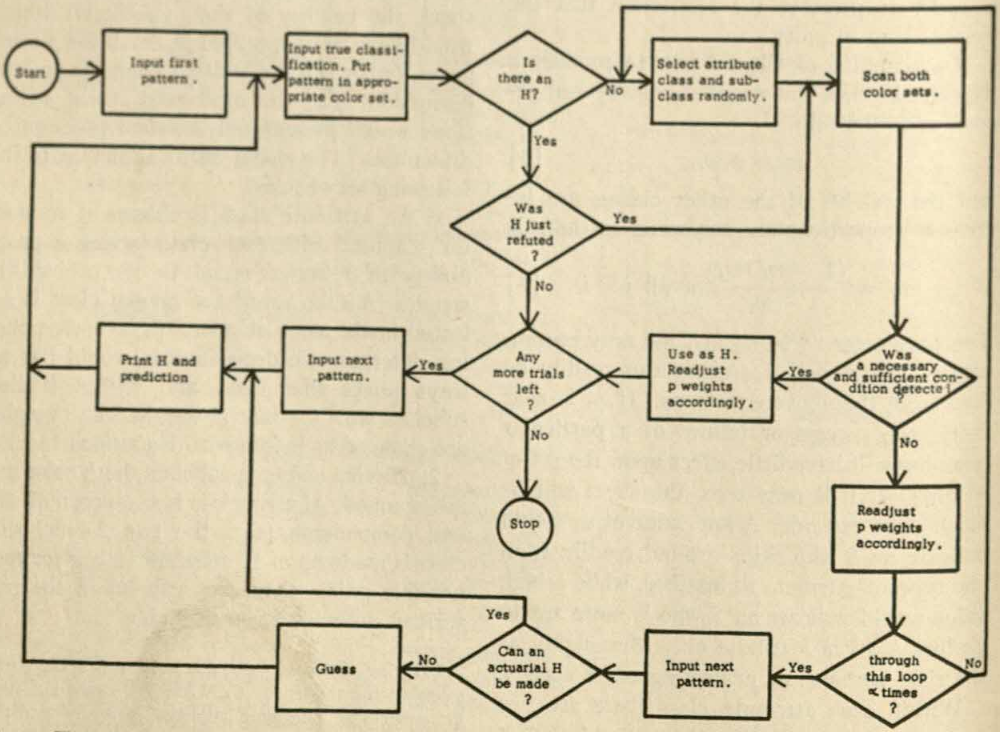

Start
Input first pattern.
Input true classification. Put pattern in appropriate color set.
Is there an H?
No
Yes
Select attribute class and subclass randomly.
Scan both color sets.
Was H just refuted ?
Yes
No
Print H and prediction
Input next pattern.
Yes
Any more trials left ?
No
Stop
Use as H. Readjust p weights accordingly.
Yes
Was a necessary and sufficient condition detecte ?
No
Readjust p weights accordingly.
Yes
No
Guess
No
Can an actuarial H be made ?
Input next pattern.
Yes
through this loop α times ?
No

FIG. 2. Flow chart for the simple model. (H stands for the current working hypothesis, p stands for probability, α is a parameter of the model.)

TABLE 7

SOLUTION TO PROBLEM 1: DETAILED EXAMPLE OF THE OPERATION OF THE SIMPLE MODEL

Line	Trial	Hypothesis or action	Prediction	Outcome	Random number	Class weights*				Subclass weights							
						Molecular	Numerical	Molar	Color	Molecular		Numerical		Molar			
										1	2	Number	Plurality	Balance	Weighting	Run	Adjacency
1	1			G		25	25	25	25	50	50	50	50	25	25	25	25
2	2	G::NB:1	G	R	32,06	12	62	12	12	50	50	75	25	25	25	25	25
3	3	Subclass rejected				23	31	23	23	50	50	38	62	25	25	25	25
4		R::B:s	G	R	67,19	12	15	62	12	50	50	38	62	62	12	12	12
5	4	Subclass rejected				22	27	31	22	50	50	38	62	31	23	23	23
6		G::W:r	R	R	51,42	11	14	65	11	50	50	38	62	15	61	12	12
7	5		G	G		11	14	65	11	50	50	38	62	15	61	12	12
8	6		G	R		11	14	65	11	50	50	38	62	15	61	12	12
9	7	Subclass rejected				21	27	32	21	50	50	38	62	27	30	22	22
10		Color cycle: GRRR	R	R	96	10	4	16	60	50	50	38	62	27	30	22	22
11	8		R	R		10	4	16	60	50	50	38	62	27	30	22	22
12	9		G	R		10	4	16	60	50	50	38	62	27	30	22	22
13	10	Color class rejected				17	25	28	30	50	50	38	62	27	30	22	22
14		Number subclass rejected			29	20	12	33	35	50	50	19	81	27	30	22	22
15		Adjacency subclass rejected			48,88	26	15	16	43	50	50	19	81	31	34	25	11
16		Actuarial: repeats are in the same color sets	R	R		26	15	16	43	50	50	19	81	31	34	25	11
17	11	Color class rejected			79	35	21	22	22	50	50	19	81	31	34	25	11
18		G::3:w	G	G	12,43	67	10	11	11	75	25	19	81	31	34	25	11

* All weights have been multiplied by 100 and rounded to two places.

The probability weights of the attrbute classes, as well as the weights of subclasses within classes, have all been set equal in the example. The parameter ϕ_1 has been set at .5, and α is 3.

In order to operate the model, a table of random numbers distributed rectangularly in the range 0–99 is used. If a choice is to be made between three alternatives with probability weights of .20, .45, and .35, say, then a random number in the range 0–19 means that the first alternative occurs, in the range 20–64 that the second alternative occurs, and in the range 65–99 that the third alternative occurs.

The model is given the information that Pattern 1 (see the problem sheet which is reproduced as Figure 1) is classified as a green. This is noted as a G. in the column headed Outcome (Line 1). The initial class weights are all set to .25. The eight rightmost columns contain the probability weights of the various subclasses. Hereafter all action of the model is determined by the operating rules, the sequence of true pattern classifications, and the sequence of random numbers, the first of which is 32. This falls in the range 25–49 and thus selects from the weights on Line 1 the class of numerical attributes for scanning. The number 06 selects the number subclass. The input to the scanner consists of a string of elements giving the numbers of black dots in successive green patterns. Since there is only one number in the string, the scanner signals that all values equal one which leads to the formation of the working hypothesis

G::NB:1.

Since a usable working hypothesis was found, the weights of the class and subclass involved are increased (Equation 1) and the other weights decreased proportionately (Equation 2). These adjusted weights appear on Line 2. The hypothesis is next used to predict that Pattern 2 will be a green. The prediction, G, is entered on Line 2 under the column headed Prediction. The outcome of the trial is found to be R so that the hypothesis is refuted. The same subclass is rescanned for another valid number hypothesis; however, no new hypothesis can be generated. This causes the rejection of class and subclass, a lowering of their weights (Equation 3), and a proportionate increase in the other weights (Equation 4). These weights are given on Line 3. The next two random numbers (67, 19) select from the weights on Line 3 the molar class and its symmetry subclass, where the valid hypothesis

R::B:s

is formed. This is rejected because of its failure to predict correctly on Trial 3. The resulting weights are given on Line 5. Although the molar-class weights are lowered (Line 5) upon the rejection of the hypothesis, they are selected again (Random Numbers 51 and 42) this time with the weighting subclass. The hypothesis found from this scanning is not refuted until Trial 6. Following this, color hypotheses are tried with success until Trial 10. On Trial 11 three subclasses are scanned without finding a valid hypothesis and so an actuarial hypothesis is used. The repeated rejection of classes and subclasses gradually raises the probability weight associated with the molecular class. When it is selected for scanning on Trial 11 the correct solution is immediately forthcoming.

Three additional solutions to Problem 1 are given in Table 8. With the exception of the

TABLE 8

THREE SOLUTIONS TO PROBLEM 1 BY THE SIMPLE MODEL[a]

Solution	Trial	Hypothesis	Prediction	Outcome	Comment
1	1			G	
	2	G::1:b	R	R	Fails on Trial 6
	7	G::3:w	R	R	Solution
2	1			G	
	2	G::NB:1	G	R	
	3	G::AW:t	R	R	Fails on Trial 5
	6	guess	R	R	
	7	G::3:w	R	R	Solution
3	1			G	
	2	G::NB:1	G	R	
	3	G::W:r	R	R	Fails on Trial 6
	7	G::3:w	R	R	Solution

[a] The initial probability weights for each of the above solutions were as follows: Molecular, .20 (1 channel, .70; 2 channels, .30); Numerical, .40 (number, .80; plurality, 20); Molar, .30 (balance, .40; weighting, .30; run length, .15; adjacency, .15); Color, .10.

initial probability weights, the same parameter values were used. Each soluton makes use of a different sequence of random numbers.

Because of negative-feedback aspects of the model (i.e., the weights associated with the "critical" attribute class and subclass tend to increase until they are selected for scanning), it is possible to specify in advance just which problems it will solve. These are: Problems 1, 2, 4, 9, 12, 22, 25, and 33. In order to be able to solve the remaining nonsequential problems, it is necessary to incorporate complex strategies into the model.

Conjunctive Strategy

This section will outline an addition to the basic model which will enable it to construct conjunctive hypotheses (i.e., hypotheses which combine two or more attribute-value pairs using the connector "and") on the basis of certain clues detected in the data. The general plan of the strategy (see Figure 3 for the flow chart) calls for a search for further necessary conditions to be combined with the first one in order to form a set which is also sufficient. The situation which triggers the use of conjunctive strategy involves a perceived necessary condition. But this is not the sole consideration, for *S* also attends to how far along he is with the problem and how close the necessary condition is to being sufficent as well.

The subroutine is entered in two ways and a parameter is associated with each. First of all, and most commonly, entry may be gained following the detection of a necessary condition. The probability of entry is

$$p = \phi_2(1 - 1/t)r, \qquad [5]$$

where ϕ_2 is a parameter in the range 0–1 which gives the limiting value of p, r is the proportion of items in the other color set not having the critical property (high values indicate that the condition is almost sufficient as well as necessary), and t is the number of the current trial. (The corresponding deterministic entry rule would set a threshold value of p above which entry is always achieved.)

Secondly, entry is possible following refutation of a hypothesis by the appearance of a pattern possessing the hypothesized properties, but belonging to the other color set. Such a refuted hypothesis may be repaired by at-

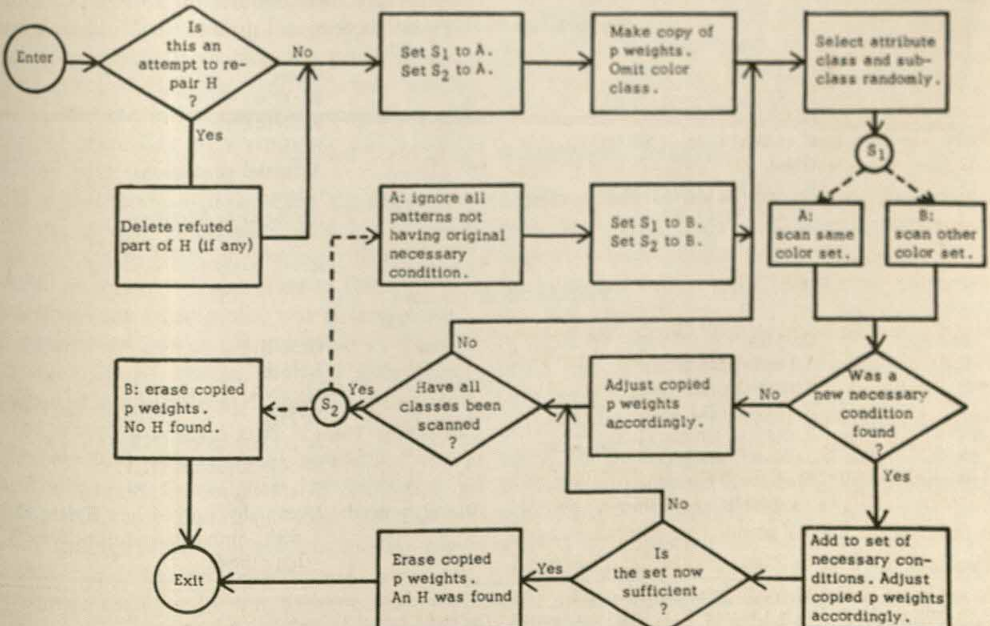

Enter
Is this an attempt to repair H ?
No
Yes
Set S1 to A. Set S2 to A.
Make copy of p weights. Omit color class.
Select attribute class and subclass randomly.
S1
A: scan same color set.
B: scan other color set.
Delete refuted part of H (if any)
A: ignore all patterns not having original necessary condition.
Set S1 to B. Set S2 to B.
No
Yes
B: erase copied p weights. No H found.
S2
Have all classes been scanned ?
Adjust copied p weights accordingly.
No
Was a new necessary condition found ?
Yes
No
Exit
Erase copied p weights. An H was found
Yes
Is the set now sufficient ?
Add to set of necessary conditions. Adjust copied p weights accordingly.

Fig. 3. Flow chart of the conjunctive-strategy routine.

taching a new necessary condition to the old hypothesis. The probability of entry under these conditions is

$$p = \phi_3(1-1/t), \qquad [6]$$

where ϕ_3 is a parameter in the range 0–1 which gives the limiting value of p, and t is the number of the current trial. (Again the deterministic rule would involve the setting of a threshold value of p.) Once a hypothesis has been repaired in this manner, further repair will always be attempted following subsequent refutations if this can be effected by the addition of another necessary condition.

There is one type of conjunctive hypothesis

TABLE 9

SOLUTIONS TO PROBLEMS 3, 5, 6, AND 20 BY THE MODEL AND BY *S*s

Model's solution		*S*'s solution	
Trial	Hypothesis	Trial	Hypothesis
Problem 3, G::(1:w)·(5:w)			
		TL	
2	R::P:4	2	Guess
3	R::AW:t	3	R::1:b
5	G::15:s	4	G::3:b
7	G::35:d	5	G::(1:w)·(5:w)
8	G::(1:w)·(5:w)	6[a]	G::15:s
		7	G::(1:w)·(5:w)
Problem 5, R:: 2 runs			
		TL	
2	G::B:s	2	G::15:s
4	Guess	5	G::3:b
5	R::A:t	6[b]	G::(NB:1) v (NB:2)
19	R::(A:t)·(15:d)	8	Vague run-length hypothesis
		10	R::[L:4] v [LB:3)·(NB:3)]
		19	R::[(L:3) v (L:4)]·[15:d]
Problem 6, R::(3:b)·(24:b)			
		SH	
2	G::15:s	2	Guess
3	G::45:s	3	R::(1:w)·(2:w)
5	G::24:s	4	R::NB:even
10	G::(24:s) unless (NB:4)	5	R::LB:2
17	Guess	6[c]	Actuarial prediction
18	R::(24:b)·(13b)·(35:b)	7	Guess
20	R::(24:b)·(3:b)	8	Actuarial predictions
			...
		15	R::(3:b)·(24:b)
Problem 20, R:: 3 runs			
		FB	
2	G::15:s	2	G::NB:1
9	Actuarial prediction	3	R::3:b
10	Actuarial prediction	5[d]	G::(NB:1) v (bwb)
11	R::(15:s)·(12:s)	7	R:15:s
18	R::(15:s) unless (L:1)	9	R::(15:s) unless (NB:3)
19	Actuarial prediction	14	R::(15:s) unless (L:1)
20 and beyond:	Model will continue to use actuarial predictions.	19[e]	R::(15:s) unless) (B:s)
		20 and beyond:	Eventually solved on Trial 50 with conjunctive-disjunctive hypothesis.

[a] An instance of generalizing an hypothesis before it is refuted.
[b] Only the black dots in Channels 2, 3, and 4 are counted. The model cannot generate such a hypothesis.
[c] Model unlikely to do this, i.e., to predict actuarially when several hypotheses can easily be found.
[d] Model unable to generate such a hypothesis. "Bwb" refers to a black-white-black sequence.
[e] Model considered same hypothesis on same trial but found it to be invalid.

which, contrary to the above equations, is often attempted quite early in a problem, namely the one-channel conjunctive hypotheses (e.g., R:: [3:b] · [4:w] · [5:b]). Consequently, whenever the scanner signals a necessary condition after scanning this subclass, an attempt will always be made to form a conjunctive hypothesis within the subclass. If this does not succeed, then entry to the conjunctive strategy subroutine will be attempted via Equation 5 in the usual manner.

Upon entry to the subroutine, the color-attribute class is temporarily deleted by apportioning its weight to the other three classes. The subroutine then conducts a search for more necessary conditions using low-level scanning. If the scanning of a subclass produces a necessary condition, its weight is increased; otherwise it is decreased. Each time a new one is discovered, it is combined with those already found to see if the set is also sufficient. If not, the search continues until at least one subclass from each of the three participating classes has been rejected. An attempt will then be made to form an "unless" hypothesis (see page 9) in the same manner. (In Figure 3, this involves the setting of both switches to B.) This procedure allows for a rather thorough search of attribute classes. Upon exiting the subroutine, the original probability weights are restored and the weights of the class and subclass which gave rise to the initial necessary condition are adjusted according to whether or not a conjunctive hypothesis was formed.

Since this model still possesses the negative-feedback aspects of the basic model, it is possible to state in advance which problems it is capable of solving. In addition to those problems which the basic model was able to solve, the present version is able to solve Problems 3, 5, 8, 17, 30, 31, and 32. Table 9 gives some results of the application of the model to some of the above problems and to one problem (Number 20) which is still insoluble. Several solutions by *S*s are given for comparison. The results show that while there are minor distinguishing features between model performance and human performance, such as hypotheses used by *S*s which cannot be generated by the model, on the whole there is little basic difference between the two types.

Disjunctive Strategy

There are a few problems (Numbers 20, 24, and 29) whose solutions require the use of disjunctive hypotheses. The subroutine for forming this type of hypothesis very closely parallels the conjunctive strategy subroutine, and the means of entry are almost identical. First, if the scanning unit signals that a particular property is a sufficient condition for one of the color sets, the probability of entry is given by Equation 5 (but where r is the proportion of items in the same color set that have the property). Secondly, if a hypothesis is invalidated by the occurrence of a pattern belonging to the same color set but lacking the hypothesized properties, then it may be repaired by adjoining a sufficient condition to the hypothesis. In this case the probability of entry is given by Equation 6.

In operation the subroutine searches for further sufficient conditions in order to find a set which is also necessary. A flow chart of the strategy is given as Figure 4. A point of dissimilarity between the two complex strategies is that the disjunctive-strategy subroutine forms nothing which is analogous to the "unless" conjunctive hypothesis.

Results of the application of the model to problems whose solutions require the use of disjunctive hypotheses are given in Table 10. Again the results are quite similar in general, if not in detail, to sample protocols.

Maladaptive Behavior

Three main types of maladaptive behavior have been identified: (*a*) learning-set phenomena which persist across problems, (*b*) the apparent failure of *S* to notice a particular type of attribute, and (*c*) failures of *S* to check hypotheses for validity. The first two types can be produced by the model merely by setting various parameters to rather extreme values. The learning set phenomenon which, for instance, seems to account for some *S*s' having attempted to solve molar problems with complex molecular hypotheses, can be created by adjusting the initial values of the probability weights in favor of the desired class of attributes. These weights can be stabilized by using a high value for the parameter ϕ_1. (Up to this point all model solutions pre-

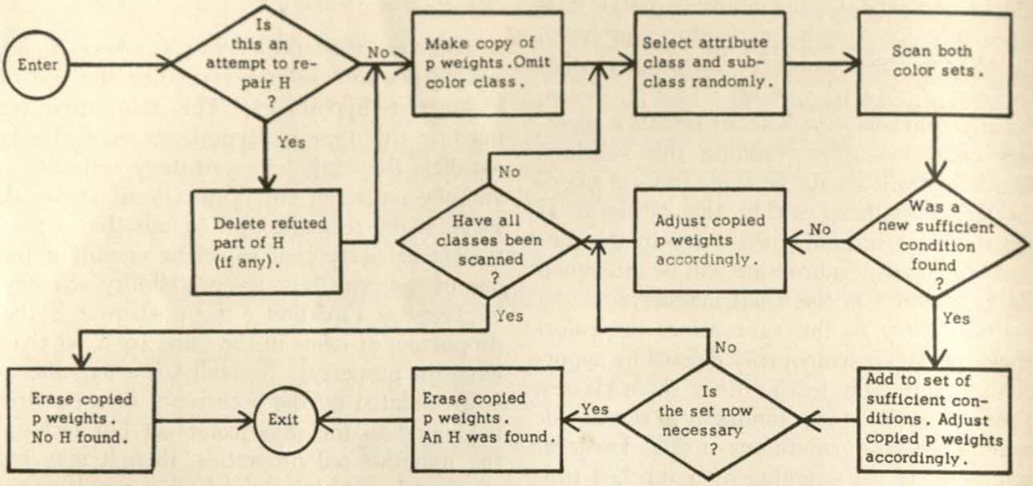

Enter
Is this an attempt to repair H ?
No
Yes
Make copy of p weights. Omit color class.
Select attribute class and subclass randomly.
Scan both color sets.
Delete refuted part of H (if any).
No
Have all classes been scanned ?
Yes
Adjust copied p weights accordingly.
No
Was a new sufficient condition found ?
Yes
Erase copied p weights. No H found.
Exit
Erase copied p weights. An H was found.
Yes
No
Is the set now necessary ?
Add to set of sufficient conditions. Adjust copied p weights accordingly.

FIG. 4. Flow chart of the disjunctive-strategy routine.

sented have used $\phi_1 = .5$, a value small enough to cause considerable variation in the weights from trial to trial.) The failure to notice a particular type of attribute can be produced simply by setting the relevant probability weight to zero initially.

The problem of accounting for *S*s' failures to validate hypotheses is more complicated. The data suggest that *S*s will occasionally perceive a necessary and sufficient conditon when, in fact, a lesser condition (necessary, sufficient, or majority) actually obtains. That is, certain invalidating cases are either overlooked or ignored. However, the greater the number of exceptions, the smaller is the likelihood that a necessary and sufficient condition will be perceived. Furthermore, when *S* is engaged in the burdensome task of trying to construct or repair complex hypotheses he is more apt to make this type of error. A function mirroring this behavior is:

$$p=\phi_4(1-r)(1-g)c,$$

where p is the probability that the results of a particular scanning will be formulated into an invalid hypothesis, ϕ_4 is a parameter giving the limiting value of p, r is the proportion of exceptions in the red color set, g is the proportion of exceptions in the green color set, and c is a nonnegative two-valued parameter: $= 1$ if the model is engaged in complex strategy; $=$ a lesser value, θ, otherwise.

This function is evaluated immediately after the scanner outputs the results of a scanning and before a possible entry to one of the complex strategy subroutines. (The corresponding deterministic decision rule would specify some threshold value of p above which the output of the scanner is always "misperceived.")

The various changes that have been made in the basic model are incorporated in the flow chart in Figure 5. Notice that it has been provided with facilities for signaling the occurrence of necessary conditions and sufficient conditions.

SIMULATION OF INDIVIDUAL SUBJECTS

In order to illustrate the behavior of the model when simulating particular individuals, three *S*s were selected who exhibited rather extreme forms of problem-solving behavior. The initial parameter values used in the model's solutions are given in Table 14.

Table 11 presents four attempted solutions to Problem 4—two by AS and two by the model. In neither of her attempts did AS reach a solution. Instead she lapsed into a series of guesses and actuarial predictions after about Trial 12. As mentioned earlier, AS seemed to behave as if there were two types of problems, one type which could be solved rather easily within 10 or 15 trials and another much more difficult type whose solution required the use of repeated, converse, and complementary patterns. In initial attempts to

TABLE 10

SOLUTIONS TO PROBLEMS 20, 24, AND 29 BY THE MODEL AND BY *S*s

Model's solution		*S*'s solution	
Trial	Hypothesis	Trial	Hypothesis
	Problem 20, R:: 3 runs		
			KC
2	G::1:b	2	G::1:b
5	Guess	5	R::(3:b)·(5:w)
6	Guess	7	R::15:s
7	G::15:d	9	Guess
9	G::(15:d) v (12:d)	10	R::(12:s)·(25:s)
18	G::(15:d) v [(12:d)·(23:d)]	18	R::[(3:b)·(1:w)·(5:w)] v [(3:b)·(1:b)·(5:b)·(2:b)]
19	G::[15:d] v [(13:s)·(24:s)]	20	· · actuarial predictions ·
		31	R::[(3:b)·(5:b)·(1:b)·(34:d)] v [(3:b)·(1:w)·(5:w)]
		32	· · actuarial predictions ·
		48	R::[(15:s)·(13:d)] v [(15:s)·(13:s)·(24:d)]
	Problem 24 R:: An isolated black dot		
			TL
2	R::15:d	2	Guess
3	R::45:s	3	R::(2:w)·(4:w)
5	R::24:s	4	R::24:s
7	R::(24:s) v (13:s)	7	R::(24:s) v (35:s)
8	R::(24:s) v (35:s)	8	G::(NB:2 v 3)·(AB:t)
10	Actuarial prediction	9	G::(NB:2 v 3 v 4)·(AB:t)
11	G::[(2:b) v (4:b)]·[AB:t]	10	G::(NB: > 1)·(AB:t)
19	G::[((2:b) v (4:b))·(AB:t)] v [NB:0]	19	G::[(NB·2v3v4)·(AB:h)] v [(12:s)·(23:s)·(34:s)·(45:s)]
20	G::[((2:b) v (4:b))·(AB:t)] v [(3:w)·(12:s)·(15:s)]		
	Problem 29 [(15:s)·(3:w)] v [(15:d)·(3:b)]		
			SH
2	R::NB:1	2	R::3:w
4	R::5:w	3	G::NB:even
7	G::(1:w)·(5:w)	4	R::(2:w)·(5:w)
8	G::(3:b)·(15:d)	6	R::5:w
14	G::[(3:b)·(15:d)] v [(15:s)·(12:d)·(NB:2)]	7	G::(23:d)·(35:d)
19	G::[(3:b)·(15:d)] v [(3:w)·(15:s)]	8	G::(3:b)·(15:d)
		14	G::[(3:b)·(15:d)] v [(3:w)·(15:s)]

simulate her behavior, (the two-channel subclass weight was set to zero to correspond to the observation that AS never mentioned an attribute of this type) it was found that if provisions were made for generating invalid hypotheses, the switch to an unbroken sequence of guesses and actuarial predictions could not take place. Instead, there was a shift to a mixture of guesses, actuarial predictions, and invalid hypotheses. Accordingly, a change was made in the model in order to bring it more into agreement with AS' apparent perception of the task. The class of actuarial relations was assigned a relatively low

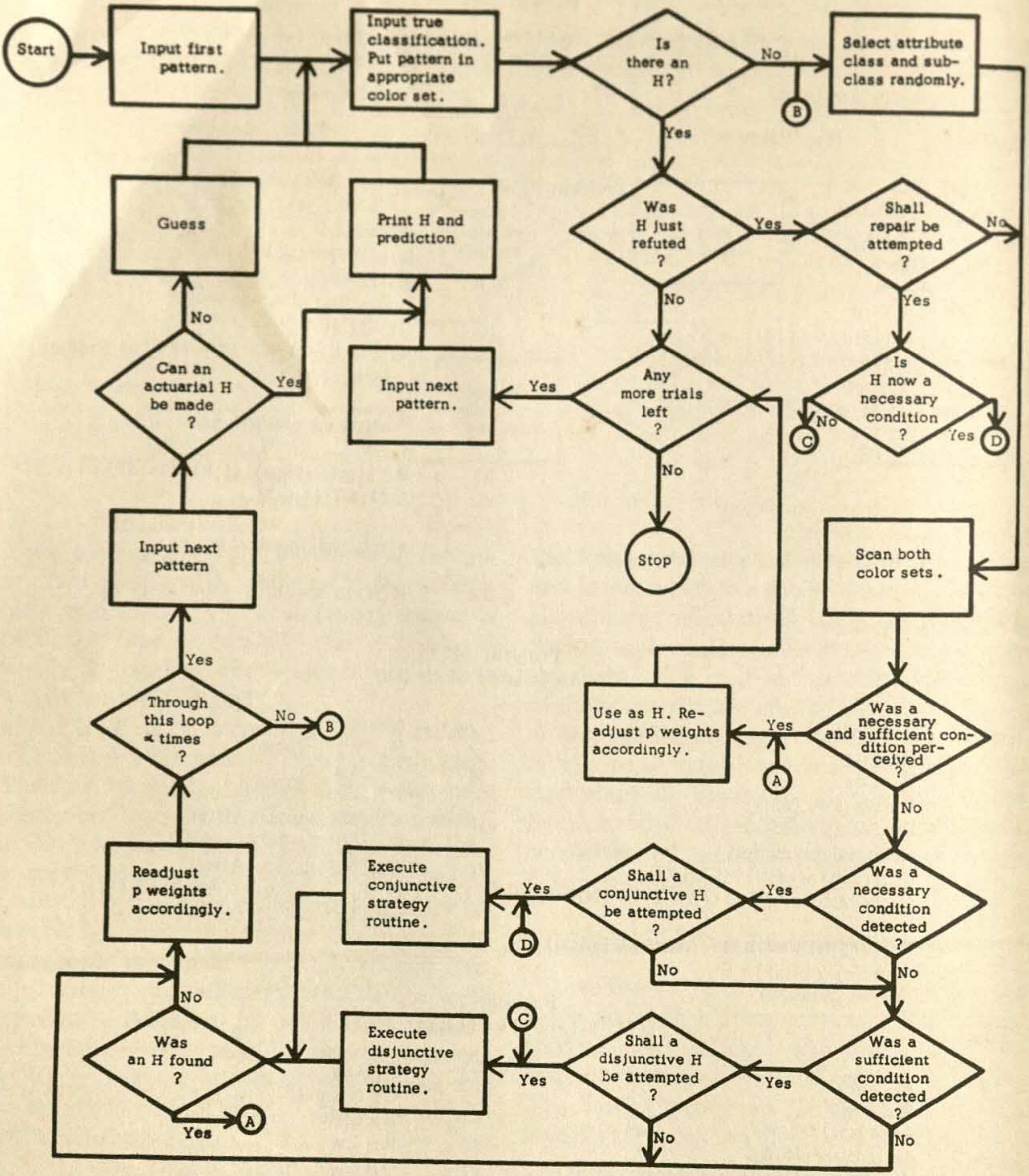

Start
Input first pattern.
Input true classification. Put pattern in appropriate color set.
Is there an H?
No
B
Select attribute class and subclass randomly.
Yes
Guess
Print H and prediction
Was H just refuted ?
Yes
Shall repair be attempted ?
No
No
Yes
No
Can an actuarial H be made ?
Yes
Input next pattern.
Yes
Any more trials left ?
Is H now a necessary condition ?
No
C
Yes
D
No
Input next pattern
Stop
Scan both color sets.
Yes
Through this loop α times ?
No
B
Use as H. Readjust p weights accordingly.
Yes
A
Was a necessary and sufficient condition perceived ?
No
Readjust p weights accordingly.
Execute conjunctive strategy routine.
Yes
D
Shall a conjunctive H be attempted ?
Yes
Was a necessary condition detected ?
No
No
No
Was an H found ?
Yes
A
C
Execute disjunctive strategy routine.
Yes
Shall a disjunctive H be attempted ?
Yes
Was a sufficient condition detected ?
No
No

FIG. 5. Flow chart of the executive part of the complete model.

probability weight and treated as if it were an attribute class. This resulted in the desired behavior. The model first searches for hypotheses in the usual way. If the solution is not forthcoming, there will come a point when hypotheses become hard to find. At this point an actuarial hypothesis is likely to be attempted because the probability weight of this class has been increasing at the expense of the weights of the other classes.

The second example, a simulation of KR's solution to Problem 24, is given in Table 12. The section on strategy mentions that KR, in her second-session solutions, used an approach which was characterized by extensive efforts to construct and repair highly complex hypotheses. The model's solution comes remarkably close to KR's solution even to generating as she did a hypothesis which is valid for 31 of the 32 patterns.

TABLE 11

SIMULATION OF AS' SOLUTIONS TO PROBLEM 4, G::25:s

Model's solution		AS' solution	
Trial	Hypothesis	Trial	Hypothesis
2	G::NB:1	2	G::NB:1
4	G::AB:t	3	G::(4:w)·(5:w)
7[a]	R::P:3	4	G::LW:2
12	And beyond: actuarial predictions	5[a]	Sequential hypothesis
		6[a]	G::LB:2 v 3
		7	Actuarial prediction
		8[a]	R::L:2
		13	And beyond: actuarial predictions
2	G::NB:1	2	Guess
4	G::5:w	3	G::5:w
6	G::(5:w)·(2:w)	6	R::AB:s
7[a]	R::B:s	7	Sequential hypothesis
9[a]	G::2:w	8	Color hypothesis
10	Actuarial prediction	10	Actuarial prediction
11[a]	R::LW: ≤2	11	(No hypothesis verbalized)
12	And beyond: actuarial predictions	12[a]	Sequential hypothesis
		14	And beyond: actuarial predictions

[a] Invalid hypothesis.

TABLE 12

SIMULATION OF KR'S SOLUTION TO PROBLEM 24, R:: ISOLATED BLACK DOT

Model's solution		KR's solution	
Trial	Hypothesis	Trial	Hypothesis
2	R::1:b	2	Guess
3	R::5:w	3	R::[(P:4)·(NB:4)] v [(P:3)·(NW:3)]
4	R::4:w	4	Actuarial prediction
6	Color hypothesis	5	R::[(3:b)·(2:w)·(4:w)] v [P:4]
7	Color hypothesis	6	R::[(32:d)·(34:d)] v [P:4]
10	R::[NB:1] v [(35:s)·(45:d)]	10[a]	R::[(32:d)·(34:d)] v [NB:1]
12[a]	R::[NB:1] v [(35:s)·(45:d)] v [(B:r)·(P:3)]	20	R::[(32:d)·(34:d)] v [NB:1] v [(NB:4)·(4:w)]
13[a]	R::[NB:1] v [(35:s)·(45:d)] v [(15:d)·(P:3)]	25	R::[NB:1] v [(NB:2)·(AB:s)] v [(NB:3)·(AB:s)] v [(NB:4)·(4:w)]
17	R::[NB:1] v [(35:s)·(45:d)] v [(AB:s)·(P:3)]	*E* Skips to trial 68,[b] which is incorrectly classified.	
E Skips to trial 68,[b] which is incorrectly classified.		69	R::[NB:1] v [(NB:2)·(AB:s)] v [(NB:3)·(AB:s)] v [(NB:4)·((2:w) v (4:w))]
69	R::[NB:1] v [(35:s)·(45:d)] v [(AB:s)·(P:3)] v [(NB:4)·(1:b)·(5:b)]		

[a] Invalid hypothesis.
[b] All the intervening trials would have been correctly classified.

TABLE 13

SIMULATION OF KC'S SOLUTIONS TO A SEQUENCE OF FIVE SIMPLE PROBLEMS

Model's solution			KC's solution		
Problem	Trial	Hypothesis	Problem	Trial	Hypothesis
1	2	G::P:4	1	2	G::P:4 v 5
	3	G::1:b		3	Color hypothesis
	7	G::3:w		4	R::12:s
				6	G::3:w
2	2	G::AB:t	2	2	G::P:4 v 5
	4	G::NB:1		6	R:NB:2
	6	R::NB:2			
8	2	G::1:b	8	2	G::NB:1
	3	G::5:w		4	G::5:w
	6	G::(2:w)·(5:w)		6	G::(2:w)·(5:w)
9	2	R::1:b	9	2	R::15:d
	5	G::(1:w)·(5:w)			
	7	Guess			
	8	G::15:s			
3	2	R::15:d	3	2	G:LB: ≦ 2
	7	R::(1:b) v (5:b)		5	G::15:s
				7	Sequential hypothesis
				8	G::(1:w)·(5:w)

The final example, a simulation of KC's solution to five consecutive problems, is given in Table 13. This sequence of solutions is characterized by a lack of maladaptive behavior and complex strategies. In fact, these solutions could have all been obtained by using the basic model specified earlier. The values of the parameters were carried over from one solution to the next.

Turing's Criterion

The last step in the presentation of a simulation model might be an experimental test of the closeness of the simulation according to Turing's criterion. In the present case a formal test of this nature would seem insured of success on a priori grounds. It can be noticed in the human solutions presented in Tables 9 through 13 that there is considerable variation in performance from *S* to *S*. Since the presented simulations seem, on the face of it, to be rather close to their human counterparts, it may be argued that the simulations fall within the wide range of individual differences. The real clues for distinguishing human from model performance seem to be the occasional hypotheses used by *S*s which cannot be generated by the model. These hypotheses, which average about one per problem, include sequential hypotheses and certain idiosyncratic

TABLE 14

INITIAL VALUES OF PARAMETERS USED IN THE SIMULATIONS PRESENTED IN TABLES 11, 12, AND 13

	Class Weight				Subclass Weight													
					Molecular		Numerical		Molar				Other parameters					
Subject	Molecular	Numerical	Molar	Color	Single	Double	Numerical	Plurality	Balance	Weighting	Run	Adjacency	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ	a
AS[a]	20	35	35	05	100	0	80	20	20	30	10	40	.7	0	0	.9	.9	3
KR	40	30	20	10	50	50	20	80	25	25	25	25	.7	.5	.9	.5	.5	3
KC	25	40	25	10	80	20	30	70	25	25	25	25	.5	0	0	0	0	3

[a] A fifth attribute class was used, actuarial, with a weight of 05.

hypotheses which were never used more than once or twice. More subtle clues, however, do exist. For instance, the model is slightly more prone to begin a problem with a series of very simple hypotheses than are *S*s. But this clue would almost certainly be overlooked by anyone not familiar with the construction of the model.

DISCUSSION

Experimental Procedure

The experiment was extremely productive of data—some 350 pages of verbatim transcripts—most of which proved to be highly valuable in helping to gain insight into the problem-solving behavior of *S*s. Several aspects of the procedure tended to insure this richness of content. First of all, the problem itself has several advantages as a concept-formation task. Within prescribed limits of format, it is possible to construct problems which cover a variety of types (the critical feature may be based upon dot position, dot arrangement, number of dots, the sequence of dots, etc.) and a wide range of difficulty. It was found in the course of the experiment that, other things being equal, more information was gained when *S*s were confronted with a fairly difficult problem. The great variety of problems available made it possible to present sufficiently challenging problems even as *S*s became more proficient at solving them.

Another important feature was the stepwise presentation of new data. There are two advantages which result from this fragmentation. First, it keeps *S* from being overwhelmed at the outset with a vast amount of data. It is quite likely that the simultaneous presentation of all the data would result either in the temporary suppression of verbal reports, while *S* engaged in fast and furious searching or in a flood of verbiage which would interfere and interact in unpredictable ways with normal searching behavior. Second, the segmentation makes natural the procedure of demanding periodic "progress reports" from *S* in the form of working hypotheses. This series of hypotheses constitutes the main point of interest in comparing *S*s' behavior with the model's behavior.

A third feature, the use of a thinking aloud procedure, was quite successful. Not a single *S* proved to be unsatisfactory in providing a running commentary of what he was doing, although many *S*s would occasionally lapse into short periods of silence while experiencing difficulty. Apparently some *S*s were able to reach a state where their reports and their thinking coincided very closely. This is seen in certain passages characterized by barely audible sentence fragments, phrases following phrases in quick succession showing a flight of ideas, etc., all spoken in low monotone broken only by the finding of some important clue. Other *S*s adopted a think-then-summarize procedure which, while not as informative about the details of their thinking, was usually adequate to specify the general line of their reasoning.

General Course of Solution

It was possible to detect among the 84 solutions a set of characteristics common to a great number of them. Almost without exception *S*s advance, on early trials, a series of extremely simple hypotheses presumed to be readily available to them. These are set forth with no great confidence that they will long survive, and little effort is made to patch them up once they are refuted. Instead their purpose seems to be one fulfilling a requirement of the experiment—produce a working hypothesis for every trial—until such a time as enough data are accumulated upon which to perform serious searching. This point is signaled, often abruptly, by a failure to spot another simple consistency.

Many problems were solved during this initial burst of simple hypotheses, and many more could have been solved. However, if a relatively easy problem had not been solved by this time, then often no solution was reached or else a roundabout solution was attained in a highly complex manner which ignored the single critical clue which was the key to the simple solution. On the other hand, if the problem was relatively hard, *S*'s success depended upon his ability to detect necessary conditions and/or sufficient conditions and put these together in the proper combination. This procedure was usually lengthy, tedious, and filled with many small errors and blunders. The proportion of invalid hypotheses rose sharply as did the number of trials on which *S* decided to "collect more data" by means of gueses and actuarial predictions.

Model

The information-processing model constructed in this study is a device for simulating certain aspects of human performance in a particular problem-solving task. Since the model is intended, ultimately, to take the form of a computer program, it is necessary to specify completely and exactly all the processes involved. For this reason the model, as with all simulation models, gives the appearance not only of extreme complexity but also of little generality. Basically, however, the model is an abstract symbol-manipulating device and, as such, should be applicable to a class of problems of which the present one is an example. Moreover there seem to be just two main characteristics of this class of problem. First, the problem must be one of finding a binary classification principle (i.e., every object to be classified must be either an exemplar or a nonexemplar). Second, all objects to be classified must be described in terms of a common set of attributes and values, and these descriptions must be made available to the model (since no artificial intelligence scheme to date is able to generate psychologically meaningful attributes *de novo*).

Any particular application of the model is apt to require the addition of a hopefully small set of special processes to handle behaviors which are characteristic of *S*s in the specific problem task. In the present case, for example, a separate routine is necessary to generate "color" hypotheses.

Implicit in the structure of the model is the notion that *S*'s problem-solving behavior is in a sense rational and optimal. The overall strategy of *S*s and model alike is to scan a set of objects for a single property (an attribute-value pair), possession of which is necessary and sufficient for membership in the set. Failing to find any such property, more complex hypotheses can be constructed on the basis of necessary conditions and sufficient conditions. What detracts from optimal performance are several human failings: invalid hypotheses are used, an attribute or class of attributes is not perceived, refuted hypotheses are persistently repaired rather than being completely rejected, complex hypotheses are avoided or used to excess. The various parameters associated with the model control the extent to which these behaviors are exhibited. This is not to say that the model is in essence an algorithm with constraints. While it is true that it can be made to operate in a manner approaching that of an algorithm, it would be more accurate to say that the model is a good problem-solving device to which have been attached several performance-reducing devices.

Once a model has been demonstrated to fit the data reasonably well, one is inclined to attempt an evaluation of the model both absolutely and with respect to alternative models. One might be interested, for instance, in comparing the present information-processing model with a mathematical SR model. Unfortunately there is considerable likelihood that such a comparison will come to grief before it can be completed. In order to understand this possibility, consider how a thoroughgoing experimentalist might carry out this monograph's concept-formation study. The only data collected in the course of the experiment would bear on the correctness of *S*'s trial-by-trial predictions. The *S* would not be asked to think aloud or even to verbalize hypotheses, but instead would be directed to keep responding until some criterion, say 10 correct predictions without an error, was met. His data would then be pooled with those of the other *S*s (all of whom respond to a different random arrangement of patterns) and a learning curve fit giving probability of a correct response as a function of trial number and problem type. As little as one parameter may be involved, and its evaluation for a particular *S* would lead to a predicted point of solution for each problem.

Thus the information-processing model and SR model would predict only one thing in common: point of solution. It is not likely that the rival sets of predictions would be very different although with a large enough N the difference could probably be made statistically significant. If the information-processing model predicts more accurately, then one can still say that SR model is nevertheless superior when accuracy as a function of number of parameters is considered, i.e., the parsimony of SR model outweighs the increased accuracy of the information-processing model. On the other hand if SR model's predictions of points of solution are superior, then one can still claim that the information-processing model, while inferior on one count, is still able to make other predictions (e.g., complexity of hypothesis) which the SR model cannot make, i.e., it is better to predict rather well on several variables than to predict quite well on just one.

More succinctly, the problem is this. Two models based upon the same set of behaviors are to be compared. One model collects more data (i.e., evaluates more parameters), operates on the data in a more complex fashion, and makes more predictions. Since this model is more difficult to construct and operate we should expect extra dividends for greater effort in terms of more extensive and more accurate predictions. But the question is how much more. How accurately should a model predict as a function of the number of parameters involved in its operation and how are models involving different numbers of parameters to be compared?

Until this theoretical question is answered a favored type of model will doubtless continue to be selected on other grounds: the comfortableness and convenience of the model's conceptual framework, the extent to which predictions are verified, the extent to which new and fruitful research is suggested, etc. Results bearing on these latter criteria will be reported at a later date.

SUMMARY

Solutions to a set of concept-formation problems were obtained individually from 11 *S*s using a thinking aloud procedure. The task required that *S* discover a principle for distinguishing between two classes of patterns. The patterns were presented sequentially, and *S* was directed to predict each pattern's classification and verbalize his current working hypothesis before being given the pattern's true classification.

Verbatim transcripts of the experimental sessions were analyzed in detail in order to discover the way in which *S*s produced the required sequence of hypotheses. A method of "low-level scanning" was identified which forms the basis for forming all simple (single attribute) hypotheses. This method involves a search for attributes common to one or the other of the two classes of patterns. The search is essentially random but with restrictions which tend to give it orderliness and direction. More complex (conjunctive and disjunctive) hypotheses may be formed upon the discovery that a particular attribute is a necessary condition or a sufficient condition for membership in one of the classes. Several points of individual difference relating to complexity of hypotheses formed, degree of validity checking, and persistence in a given line of attack were identified.

An information-processing model was constructed which is capable of generating a sequence of hypotheses in response to the presentation of a sequence of patterns. The model's style of solution is determined by preadjusting various parameters relating to individual differences. Several solutions by the model are presented which closely parallel selected human solutions. Overall, the model was judged by the author to be an adequate simulation of human behavior in the given concept-formation situation.

REFERENCES

BARTLETT, F. *Thinking: An experimental and social study*. New York: Basic Books, 1958.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1957.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.

DUNCKER, K. On problem solving. *Psychol. Monogr.*, 1945, **58**(5, Whole No. 270).

FEIGENBAUM, E. The simulation of verbal learning behavior. *Proc. West. Joint Comput. Conf.*, 1961, **19**, 121–132.

FELDMAN, J. Simulation of behavior in a binary choice experiment. *Proc. West. Joint Comput. Conf.*, 1961, **19**, 133–144.

GAGNÉ, R. M. Problem solving and thinking. *Annu. Rev. Psychol.*, 1959, **10**, 147–172.

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59**, 461–472.

HULL, C. L. Quantitative aspects of the evolution of concepts: An experimental study. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

HUNT, E. B. *Concept learning: An information processing problem.* New York: Wiley, 1962.

JOHNSON, D. M. Problem solving and thinking. *Annu. Rev. Psychol.*, 1950, **1**, 297–310.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

JOHNSON, E. S. The simulation of human problem solving from an empirically derived model. Unpublished doctoral dissertation, University of North Carolina, 1961.

JOHNSON, E. S. Simulated problem solving: Summary of a model and some representative data. Research Memorandum 10, January 1962, Psychometric Laboratory, University of North Carolina.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Empirical explorations of the logic theory machine. *Proc. West. Joint Comput. Conf.*, 1957, **11**, 218–230.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Elements of a theory of human problem solving. *Psychol. Rev.*, 1958, **65**, 151–166.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Report on a general problem-solving program. In, *Proceedings of the First International Conference on Information Processing.* Paris: Unesco, 1959.

NEWELL, A., SHAW, J. C., & SIMON, H. A. A variety of intelligent learning in a general problem solver. In M. T. Yovitts, & S. Cameron (Eds.), *Self-organizing systems,* New York: Pergamon Press, 1960.

SCHEERER, M. Cognitive theory. In G. Lindsey (Ed.), *Handbook of social psychology.* Vol. 1. Reading, Mass.: Addison-Wesley, 1954. Pp. 91–142.

TAYLOR, D. W., & MCNEMAR, OLGA W. Problem solving and thinking. *Annu. Rev. Psychol.*, 1955, **6**, 455–482.

TURING, A. M. Can a machine think? In J. R. Newman (Ed.), *The world of mathematics.* New York: Simon & Schuster, 1956. Pp. 2099–2123.

VINACKE, W. E. *The psychology of thinking.* New York: McGraw-Hill, 1952.

WANG, H. Toward mechanical mathematics. *IBM J. Res. Develpm.*, 1960, **4**, 2–22.

WERTHEIMER, M. *Productive thinking.* New York: Harper, 1945.

(Received August 30, 1963)

APPENDIX

Verbatim Transcript

Periods of silence are indicated by an ellipsis (. . .). A silent period of 3 or 4 seconds is long enough to be so denoted. Thereafter another ellipsis is added for every 10–15 seconds of silence.

For further verbatim transcripts see Johnson (1962).

E: This will be Problem 29. The time is 2:23 and Number 1 is a red.

S: Well, call Number 2 green. Because the center dot is the opposite color from the first one.

E: Number 2 is red.

S: Let's see. I call Number 3 green because it has an even number of black dots.

E: Number 3 is red.

S: Well, now the last three have in common the fact that the second dot is white and that the fifth dot is white, so I'll use both of 'em the first time. Number 4 has go the second dot white and the last dot black so I'll say that that's green.

E: Right.

S: That'll make five green too.

E: Five is really red.

S: O.K., five is red . . . obvious—the first thing that strikes me is that the red ones have in common is that the fifth dot is always white. So six would be green.

E: Six is red.

S: Oh, Lord . . . Seems I always stump on about that sixth one and start looking back. Uh . . . The green has the middle dot and the fifth dot both black but so does the sixth one. I suppose there's the possibility that there's a three-dot combination but that seems a little too complicated to work with . . . Let me see . . . if first dot, first, third, and fifth, that's not white . . . That might work. I'm going to try and see if the first dot is an opposite color from the third and the fifth dot. So that would make seven, wait, no, that wouldn't work because Number 5 is red. The second dot opposite from the third and fifth dot . . . I'll try that. Number 7 is red.

E: 7 is green.

S: . . . The two green ones have in common the fact that the middle ones are black and also the fourth ones are white . . . but the red ones have that in common too, or some of them do. Ah . . . I'm thinking about that tricky one I missed last time now where the coloring of the black dots in the previous distribution determined the color of the next one some how or other. I've forgotten how it worked. Let's see. I can't even think that way. I'm going to forget about it. Ah . . . . . . I see something here. It's another one of these wild things. Seems that on the green ones if the middle one is black, the first one and the 5th one are opposite in color. Now that doesn't show up in any of the red ones, I don't think it's right, but I'm going to try it if I can remember it. Number 8 would be red.

E: Right.

S: Number 9 is green.

E: Right.

S: Number 10 is green.

E: Right.

S: Number 11 is red.

E: Right.

S: Number 12 is green.

E: Right.

S: Number 13 is red.

E: 13 is green.

S: Mm . . . . O.K., I'm going to try adding something to it then. If the middle dot is black, the first and fifth dots will be opposite in color. If the middle dot is white the first and fifth will be the same in color . . . 14 is red.

E: Right.

S: 15 is green.

E: Right.

S: 16 is red.

E: Right.

S: 17 is red.

E: Right.

S: 18 is, is, is, I've forgotten what I was doing. 18 is green. I'm sure.

E: Right.

S: But I can't remember what I'm doing.

E: —

S: Un-huh. I can't formulate it in my mind, you know? 19 is red.

E: 19 is really green.

S: I did forget it then. All right. If the center . . . yea, if the center dot is black . . . well, what in the world . . . O.K., if the center dot was black, they were supposed to be opposite in color. If it was white, they were supposed to be the same in color. That's right. That one was green, then, wasn't it? 19's green?

E: Right.

[*S* continues to predict correctly for several more trials.]

E: O.K., what is it?

S: That's a messy one. I still have to look at it. If the center dot is white, the first and fifth dots will be the same color as each other. Not necessarily the same as the center dot. And the center dot is black, the first and fifth will be opposite color from each other. I can't keep 'em straight.

CONTENTS

Cue Probabilism and Inference Behavior 1

Lee Roy Beach

Recognition, Assimilation, and Identification of Objects 21

Lee Roy Beach

Vol. 78, No. 5 Whole No. 582, 1964

Psychological Monographs: General and Applied

CUE PROBABILISM AND INFERENCE BEHAVIOR[1]

LEE ROY BEACH

Office of Naval Research, Washington, D.C.

This is an investigation of how Ss use probabilistic cues to make inferences about objects' (Os') class memberships. Os having specified probabilistic cue properties were presented and Ss' inferences were compared to the theoretical predictions. It was concluded that: (a) for unfamiliar Os cue probabilism is important in class inferences; (b) for familiar Os inferences are based upon recognition; (c) for unfamiliar Os which are similar to familiar Os the inference is the class appropriate to the familiar O; (d) Ss who rely upon recognition also rely upon the similarity between familiar and unfamiliar Os in order to make inferences for unfamiliar Os.

Brunswik's (1956) theory of probabilistic functionalism presents not only a point from which to view the behavior of men in their environment, but also provides a framework within which further theorizing can take place. The present work is an attempt to use this theory as a point of departure for an investigation of how subjects (*S*s) make inferential judgments about the identity of objects (*O*s) encountered in the environment.

Probabilistic functionalism is a theory of perception and cognition. It views *S*s as operating in a complex environment in which cues associated with *O*s are not perfectly correlated with the attribute being inferred. For example, height correlates approximately .10 with IQ and is therefore a poor cue upon which to base inferences about IQ. Vertical position in the visual field, however, correlates .60 with an *O*'s distance from *S* and thus is a moderately good cue for inferring distance (Brunswik, 1956).

The equivocal relationship between cues and *O*s, called cue probabilism, requires that *S*s learn to utilize optimally the various kinds of cues. Correlational analyses of inference behavior show that the correlations between each of the different kinds of cues and *O*s are paralleled by the correlations between *S*s' inferences and the cues. Moreover, the validity of *S*s' inferences about *O*s (also measured by correlation) is higher than the correlations between any one of the different kinds of cues and *O*s (Brunswik, 1956; Rappoport, 1961; Smedslund, 1955; Summers, 1962). These findings indicate that *S*s learn to utilize probabilistic cues and that by combining the information from multiple cues they can make accurate inferences about *O*s.

Because the theory of probabilistic functionalism is stated in terms of correlation it is not surprising that the related research has relied upon this analytic technique. The advantage in the use of correlation is, of course, that it permits succinct state-

---

[1] Opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or the endorsement of the Navy Department.

The first part of this investigation was conducted while the author was associated with the Behavioral Research Laboratory at the University of Colorado and was supported by Public Health Service Fellowship MF-12, 744, from the National Institutes of Health. The second experiment was conducted at the United States Naval School of Aviation Medicine, Pensacola, Florida.

The author wishes to express his sincere appreciation to R. K. Hammond, M. Wertheimer, J. R. Berkshire, Alice Garman, and Evelyn Helvey for their valuable help and advice.

ments about *Ss'* use of cues for inferences about a group of *Os* or for repeated inferences about *O*. The disadvantage is, however, that the unique event—a single inference for one *O*—is glossed over in favor of the more general correlational statement. It would be valuable to have an alternative theoretical and analytical approach to inference which would permit the examination of the inferences made for each *O* as a unique event. The present work attempts to provide such an approach.

Nature of the Environment

Cues. The first assumption is that the immediately apparent, palpable characteristics of an *O* are the bases for inferences about its covert, impalpable attributes. For example, an *O's* size, surface texture, weight, color, and shape all supply information about what kind of *O* it is—whether *O* is a house, a cup, a peach, etc. The immediately apparent characteristics of *Os* are called *cues*.

Cue Dimensions. If *Os* are compared for some characteristic, such as weight, it is clear that different *Os* possess the characteristic to different degrees. What is more, if a large number of *Os* are compared on this characteristic, it is possible to order them in some manner. The metric created by the ordering of *Os* according to their possession of an attribute is called a cue dimension. The value on a cue dimension which corresponds to the degree to which an *O* possesses the related characteristic is called *O's* cue value. (Dichotomous cue dimensions are considered as special cases of the more general multivalued dimension.)

Covert Attributes. Os' cue values lie at one pole of the inference process. At the other pole is the goal of the inference—the covert, not immediately apparent characteristics of *O*. The covert attributes are of two kinds, *O's* identity, its class membership, and other attributes of *O* which *S* does not know, its covert cue values.[2] The covert attribute with which the present work is concerned is *O's class membership*. The inference process goes from the (overt) cues associated with *O* to the identification of the (covert) class to which *O* belongs. In this situation *S* must rely on the cues, make an inference about the class, and finally receive information of some kind about the accuracy of his inference.

Many definitions of the term class have been used in previous research (Bruner, Goodnow, & Austin, 1956; Sarbin, Taft, & Baily, 1960). The definition to be used here is that a *class* is a name assigned to two or more *Os*. This name is used to identify *Os*, and its use is usually dictated by a source external to *S*. Indeed, the process of a child learning about *Os* in his world consists in part of forming links between different cues and the names which elders indicate to be "correct."

Probabilism of Cues. The process of inferring an *O's* class from its cue values is complicated by the fact that a given cue value has usually been associated with a number of different classes of *Os*. Consequently, knowledge of this value does not allow an unequivocal inference about *O's* class membership. This equivocality, which is inherent in the use of cues as the bases of inferences, is called *cue probabilism*.

Cue Probabilism and Classes. In the environment every *O* possesses cue values and also belongs to a class. Theoretically, for each cue value on each cue dimension it is possible to construct a distribution of the relative frequency of association of the cue value with various classes of *Os*. If it is assumed that events which have occurred with the greatest relative frequency in the past are the events most likely to occur in the future, this distribution can be used to make inferences about the class membership of *Os*. Given a cue value associated with an *O*, the class having the greatest relative frequency of association with the cue value is most likely to be the correct class for *O*.

The relative frequency distributions for cue-class associations are called *cue distributions* and are the keystone of the present approach to inference.

Nature of the Subject

Cues. Basically the definition of cues must rely on psychophysically defined units along dimensions of physical stimulation; however, it is obvious that with experience "higher order" dimensions are developed. For example, numbers, while abstract, convey a great deal of information about *O*s with which they are associated. Experience not only leads to higher order cue dimensions but, in addition, individual differences in experiential background may produce individual differences in the dimensions to which *S*s attend and individual differences in the size of the units on the dimensions. Studies by Beach and Wertheimer (1961); Brunswik (1956); Goodnow (1954); Hays (1958); Osgood, Suci and Tannenbaum (1957); Signell (1961); Wertheimer (1960); and others demonstrate these individual differences. While these variables must be examined when research is done, for simplicity of discussion it is assumed momentarily that *S*s are attending to all dimensions and that the sizes of the cue units on these dimensions are the same for all *S*s.

---

[2] While for the purposes of the present discussion it can be assumed that fundamentally there is no difference between the cues and the covert attributes, the fact that cues are observable at the time the inference is made while the covert attributes of *O* are not, permits a psychologically meaningful distinction between them.

Classes. The cue distributions have a subjective counterpart resulting from *S*'s experience with *O*s and their classes. It is assumed that during this experience *S* compiles the relative frequencies of class occurrence for the values on the cue dimension and, as a result, knows the distribution of cue-class associations for each cue value on each cue dimension in the environment. These subjective counterparts of the environmental distributions are called *S*s' *distribution hypotheses* (Brunswik, 1956).

The assumption of distribution hypotheses regards *S* as an actuary (Meehl, 1954; Sarbin, 1943). And, just as an actuary's tables become more accurate with an increased sample size, so too do distribution hypotheses become more like their objective counterparts as *S*'s experience with various cues and various classes of *O*s increases. With increased experience, *S*'s distribution hypotheses should become highly accurate representations of the corresponding environmental cue distributions. This point will assume a great deal of methodological importance in the experiments to be reported.

In addition to the differences in attention to cues, it should be noted that *S*s' distribution hypotheses may not conform to their objective counterparts. While little is known about the determinants of the discrepancies it is likely that as an *S*'s "sample size" increases, his distribution hypotheses approach accuracy (see Attneave, 1953; Edwards, 1961; Preston & Baratta, 1948, for related research). For the sake of simplicity in the following discussion it is assumed that *S*s' various distribution hypotheses are accurate.

Nature of the Inference Process

When an *O* is encountered in the environment, the cue values associated with it serve as the basis for a decision about its class membership. To make this decision *S* must rely upon his distribution hypotheses. Due to the fact that any cue value has usually been associated with more than one class of *O*s, the cue does not allow an unequivocal decision about the class membership of the *O* which possesses it. However, for each cue value associated with *O*, the related distribution hypothesis tells *S* the relative frequency with which the cue previously has been associated with the members of each class. The *O* should be assigned to the class for which the relative frequency is highest. Of course, while one cue may indicate a specific class to be the most likely alternative for *O*'s membership, other cues may indicate other classes to be most likely.

When an *O*'s various cues do not yield an unequivocal inference, the best that *S* can do is to gather together the information provided by each of the cues and to arrive at a compromise which will allow him to determine which class is the "best bet" (Brunswik, 1943) for *O*'s class membership. This best bet should be *S*'s inference for that *O*. The following discussion describes a simple method for using the probabilistic aspects of cues to arrive at the best bet for an *O*'s class membership and for predicting *S*s' inferences.

Inference Model. The probabilistic relationships between cues and classes are the key to the use of an *O*'s cue values to determine the best bet for its class membership. When an *O* is encountered in the environment, some of its cue values are immediately known. However, *O*'s cue value on each dimension *could* be associated with a member of any of a number of different classes. As a result, the *O* possessing the cue values *could* be a member of any one of these classes. Therefore, it is necessary to compile, for each class, the evidence which indicates that *O* belongs to that class. Then the evidence for each class must be considered, relative to the evidence for each of the other classes, and a decision made about which class is the best bet for *O*'s membership. Finally, *O* should be relegated to the class which was determined to be the best bet.

Equation 1 describes a simple way to "compile the evidence" for each class.

$$\sum_{d=1}^{n} P(k/d) = P(k/1) + P(k/2) + P(k/3) \cdots P(k/n). \qquad [1]$$

Where,

k = a specific class under consideration as a possibility for *O*'s membership; k = a through i

d = a dimension on which *O*'s cue value is known; d = 1 through n

$P(k/d)$ is the relative frequency with which *O*'s cue value on each cue dimension, d, has been associated with members of a given class, k.

Equation 2 describes a way to compare the "relative amounts of evidence" for the different classes.

$$E(k) = \frac{\sum_{d=1}^{n} P(k/d)}{\sum_{k=a}^{i} \sum_{d=1}^{n} P(k/d)}. \qquad [2]$$

Where,

$\sum_{d=1}^{n} P(k/d)$ is the value for one class, k, derived from Equation 1 and

$\sum_{k=a}^{i} \sum_{d=1}^{n} P(k/d)$ is the sum of these values over all classes, a through i.

The denominator of Equation 2 can be simplified. The proportion of times a cue has been associated with each class sums to 1.00 over all classes; a cue never occurs without an *O*, and *O*s always belong to classes. Therefore, the denominator of Equation 2 becomes,

$$\sum_{k=a}^{i} \sum_{d=1}^{n} P(k/d) = 1.00 \times n = n.$$

Where,

n is the total number of cue dimensions for which *O*'s cue values are known.

Replacing the denominator in Equation 2 with n yields Equation 3.

$$E(k) = \frac{\sum_{d=1}^{n} P(k/d)}{n}. \qquad [3]$$

Equation 3 gives the proportion of the total evidence which indicates that *O* is a member of Class k. This, $E(k)$, can be regarded as the degree to which a given class can be expected to be correct in light of *O*'s immediately apparent cue values. The best bet for *O*'s class membership is the class having the highest $E(k)$ as given by Equation 3.

METHOD

Theoretically, *S*s' inferences about *O*s are predicated upon the statistical structure of the environment. Therefore to test the utility of the model an experimental situation was constructed which had controllable statistical characteristics and then, because distribution hypotheses are assumed to develop with experience, *S*s were allowed experience with the "experimental environment" sufficient to develop accurate distribution hypotheses. Finally, *S*s' inferences for specific *O*s were compared to the inferences which were predicted by the model.

To compare *S*s' inferences with those predicted by the model, data were obtained from two sources: (*a*) inferences made for specially designed test *O*s and (*b*) inferences made for each of the *O*s which comprised the experimental environment. The latter are called the environmental *O*s.

Test Os

The use of test *O*s required that *S*s first be given enough experience in the experimental environment to develop accurate distribution hypotheses. Then test *O*s were presented, and the model's decision about the best bet for each test *O*'s class was compared with the *S*s' inferences for the test *O*. This procedure allowed an estimation of the utility of the model for predicting *S*s' inferences about an *O* which has never before been encountered.

Environmental Os

Inferential Accuracy. Our definition of a class as a covert attribute about which *S* does not know provides a criterion for designating an inference as either accurate or inaccurate. An inference is inaccurate when an *O* is assigned to a class which, in the light of later information, proves to be incorrect.

Inferential accuracy was an important tool for the investigation of *Ss'* inferences. To evaluate the utility of the model for predicting inferences for test *Os*, it was sufficient to compare *Ss'* inferences for each test *O* with the theoretical best bet for that test *O*; *Ss* either chose the same class as the model or they did not. For the environmental *Os*, however, it was possible to examine *Ss* inferences for *Os* repeatedly encountered in the experimental environment. Here the course of *Ss'* increase in inferential accuracy could be examined in light of our assumptions about the inference process. It was predicted that in the early stages of training the accuracy for *O* would be low; however, as experience increased and the distribution hypotheses became more accurate, the accuracy of inferences for the given *O* would increase.

At first glance it appears as though an increase in inferential accuracy could be accounted for merely by paired-associates learning of *Os* and their classes. However, by using the model it was possible to specify *Os* for which accuracy should not increase, as well as the errors which should be made for these *Os*. If there were no differences between *Ss'* inferential accuracy for these *Os* and their inferences, the present theory could be regarded as incorrect. If, however, *Ss'* inferential accuracy for these *Os* differed from the accuracy for other *Os*, it would constitute strong support for the validity of the present approach to inference behavior.

Confirming and Infirming Cases. When the model was used to infer the class membership of an *O* and the inferred class is the one to which it actually belongs, it was called a *confirming* case. On the other hand, some *Os'* cue values lie in the tails of the cue distributions. Applying the model to these *Os* resulted in designating an incorrect class as the best bet for their class membership. These *Os* were called *infirming* cases.

It was predicted that if the model is an accurate analogy to *Ss'* inference behavior, inferential accuracy for confirming cases should increase as experience in the environment increases, i.e., as distribution hypotheses become accurate. On the other hand, *Ss'* inferences for infirming cases should follow the errors of the model. These cases should be assigned repeatedly to the class which the model dictates to be the best bet even though it is incorrect. Thus, inferential accuracy for infirming cases should decrease as *Ss'* distribution hypotheses become accurate, and the class to which they are assigned should be the model-determined best bet. In short, *Ss* should make the same inferences as are made by the model.

In summary, if the theory is correct and the model is adequate, it was predicted (*a*) that *Ss* would learn to use the probabilistic aspects of the cues associated with *Os* in order to make accurate inferences about *Os'* class memberships; (*b*) that for test *Os* *Ss'* inferences would correspond to the class designated as the best bet by the model; (*c*) that for environmental *Os* which were identified by the model as confirming cases *Ss'* inferential accuracy would increase as experience in the experimental environment increased; (*d*) that for environmental *Os* which were identified by the model as infirming cases *Ss'* inferential accuracy would decrease as experience in the experimental environment increased; (*e*) that *Ss'* inference errors for infirming cases would result from assigning these *Os* to the class corresponding to the class designated as the best bet by the model.

EXPERIMENT I

Construction of the Experimental Environment

The experimental environment consisted of a deck of 3 × 5 inch cards, each of which was considered an *O*. On its face each card possessed one value from each of three cue dimensions: (*a*) letter (Values *a* through *i*), (*b*) number (Values 1 through 9), and (*c*) pointer position (Values 1 through 9 in a clockwise direction). The class to which each card belonged—red, yellow, or blue—was written on its back.

In order to specify the sets of cues and the class for each card, arbitrary frequency distributions were defined for each class on each cue dimension. These distributions, presented in Table 1, were then used as the marginals on three 9 × 9 × 9 matrices, one matrix for each of the three classes. The marginal distributions determined frequencies in the matrix cells, and each matrix cell represented a combination of three cue values. Next, cards bearing each cell's cue combination were made; the number of cards made was dictated by the frequency (never larger than six) in the particular cell, and the color on their backs was dictated by the matrix to which the cell belonged. This procedure yielded a deck of 120 cards, 40 in each class.

There were 62 different cue combinations. (In the remainder of the discussion, individual cards will be referred to according to their cue combinations. For example, f26 means a card possessing cue value f from the letter dimension, cue value 2 from the number dimension and the cue value 6 from the pointer dimension. There were four cards bearing the cues f26 in the deck.) All three cue dimensions had approximately equal correlations with the classes (contingency coefficients of .51 for letter, .51 for number, .40 for pointer).

After the deck was constructed $P(k/d)$ was computed for each value on each dimension for each class. These $P(k/d)$, presented in Table 1, were then used in the application of the model to each card in the deck in order to obtain the theoretical estimate of the class which was the best bet for the card. Ordinarily the class with the highest value of $E(k)$ was the class to which the card actually belonged; when this occurred, the card was designated a confirming card. For some cards, however, the model yielded values of $E(k)$ which indicated that the best bet for their class membership was a class other than the one to which they

TABLE 1

FREQUENCY DISTRIBUTIONS FOR CUE VALUES' ASSOCIATION WITH EACH CLASS AND THE CORRESPONDING VALUES OF $P(k/d)$: EXPERIMENT I

Class	Letter								
	a	b	c	d	e	f	g	h	i
Red									
f	0	0	0	0	4	6	10	14	6
$P(k/d)$	.00	.00	.00	.00	.20	.43	.71	.93	1.00
Yellow									
f	0	1	4	8	14	8	4	1	0
$P(k/d)$	.00	.07	.29	.57	.70	.57	.29	.07	.00
Blue									
f	8	14	10	6	2	0	0	0	0
$P(k/d)$	1.00	.93	.71	.43	.10	.00	.00	.00	.00
	Number								
	1	2	3	4	5	6	7	8	9
Red									
f	0	1	4	8	14	8	4	1	0
$P(k/d)$	.00	.07	.29	.57	.74	.53	.31	.08	.00
Yellow									
f	4	14	10	6	4	2	0	0	0
$P(k/d)$	1.00	.93	.71	.43	.21	.13	.00	.00	.00
Blue									
f	0	0	0	0	1	5	9	11	14
$P(k/d)$	.00	.00	.00	.00	.05	.33	.69	.92	1.00
	Pointer								
	1	2	3	4	5	6	7	8	9
Red									
f	4	14	10	6	4	2	0	0	0
$P(k/d)$	1.00	.93	.71	.43	.18	.12	.00	.00	.00
Yellow									
f	0	0	0	0	4	6	10	14	6
$P(k/d)$	.00	.00	.00	.00	.18	.38	.71	.93	1.00
Blue									
f	0	1	4	8	14	8	4	1	0
$P(k/d)$	.00	.07	.29	.57	.64	.50	.29	.07	.00

actually belonged. These cards were designated infirming cards.[3] This procedure resulted in 53 different kinds of confirming cards and 9 different kinds of infirming cards.

In addition to the deck of cards, a number of test cards (test *O*s) was made. There were 4 test cards having three cue values on them and 10 having two cue values on them. The combination of cues on these cards was unlike any of the combinations on the cards in the deck. Three blank cards were also included among the test cards.

Procedure

The *S*s were 40 undergraduate university students who met as a group during the time they were gaining experience with the experimental environment. As will be explained below, only 10 *S*s were used in the data analysis. On the first day it was explained to *S*s that by attending to the cues on the front of the cards they could come to make accurate inferences about the class to which the card belonged. They were further told, "each card is an individual object and is to be thought of and treated as such." The procedure was then explained, and *S*s were told to look at the cues on the face of the card, make a decision about the card's class, and write an inference on the answer sheet. After a 4–5 second delay the experimenter (*E*) would reveal the correct class for that card and present the next card. If the inference was correct, *S*s were instructed to place a check beside it on the answer sheet. This was done to increase *S*s' involvement in the task.

---

[3] It should be understood that infirming cases were not violations of the statistical properties of the experimental environment. Instead, they were, in a sense, errors made by the model. An infirming case appears a certain proportion of the time relative to the other *O*s in the environment, and its repeated occurrence does not lead to an increase in the relative freqeuncy of its values in the cue distribution for its class. The values used in the model are the relative frequencies from these distributions and, therefore, each time the infirming case appears, the same relative frequencies are used in the computations. In this experimental situation, given the same cue values, the model will always assign the infirming case to the incorrect class.

Twelve identical decks of the 120 cards were each shuffled and placed one after another in a box. Although *S*s were unaware of where one deck left off and another began, they were shown the first 2 of the 12 decks in the first training session. Decks 3, 4, and 5 were seen in the second session; Decks 6, 7, and 8 in the third session; and Decks 9, 10, and 11 in the fourth session. Deck 12 was seen later, in a fifth session. Each time a full deck was seen will be referred to as a trial. For the fifth training session, *S*s met individually with *E* and were shown the twelfth deck (Trial 12). Then, *S* was shown the test cards and was asked to make a decision about their class membership.

Results

Cue Dimension Use. In order for the computations for the model to be applicable in predicting *S*s' responses, it was necessary to determine which *S*s were relying on all three dimensions. To do this it was assumed that *S*s' distribution hypotheses would be most accurate for the dimensions to which they attended. Therefore, the group was shown a blank test card and asked what cue value from each of the dimensions would be most likely to appear on the card's face if it were known that the card was, say, a member of the red class. Then, for each *S*, the absolute difference was obtained between his judgment and the actual mode of the related cue distribution. When the mean of these differences for one of the cue dimensions was at least twice as large as the mean of the differences for the other two dimensions, it was assumed that *S* was not attending to the dimension. Only the 10 *S*s found to be using all three dimensions were used in the data analysis.[4]

Inferential Accuracy. The first prediction was that *S*s would be able to use the probabilistic nature of cues to make accurate inferences about *O*s. The curve in Figure 1 shows the mean number of inferential errors for each of the 12 trials and clearly shows an increase in inferential accuracy accompanying an increase in experience. The decrease in errors as a function of experience in the experimental environment (the training deck), together with the rest of the data to be presented, are taken as support for the prediction.

Test Cards. The second prediction was that *S*s' inferences for the class membership of test cards would correspond to the best bet as dictated by the model. The results for this prediction are shown in Table 2.

This table is divided into 14 horizontal sections, one for each test card. In each section there are 5 columns. The first column contains the cues which were on the test card. (A dash signifies the absent cue on the test cards which have only two cue values.) The second column indicates the three classes, and the third contains the values of $E(k)$ corresponding to each class. The class and value of $E(k)$ set in boldface type in the second and third columns designate the best bet on the basis of the model. The fourth column presents the number of *S*s who gave each class as their inference for the test card. The last column gives the value and significance for a one-sample chi-square test for the results in the fourth column.

The response distributions for 10 of the 14 test cards differ significantly from chance. For all of these 10 the greatest number of *S*s chose the class which the model dictated to be the best bet. Of the four test cards yielding insignificant response distributions one, c-6, is in the predicted direction and another, d54, shows a split response distribution corresponding closely to the nearly equal values of $E(k)$ for classes red and blue. These results support the hypothesis

---

[4] Because it may appear that the results to be reported are due to selection of *S*s it should be noted that a similar analysis for the 20 *S*s who were using only the letter and pointer dimensions yielded virtually identical results. The two-cue analysis required deletion of the number dimension from the computations for every $E(k)$ in the experiment.

Because the experiment was originally designed for three-cue dimensions the two-cue theoretical predictions were frequently not as clear-cut as the three-cue predictions. Therefore, to avoid needless repetition and complexity in presentation the two-cue results are not presented here.

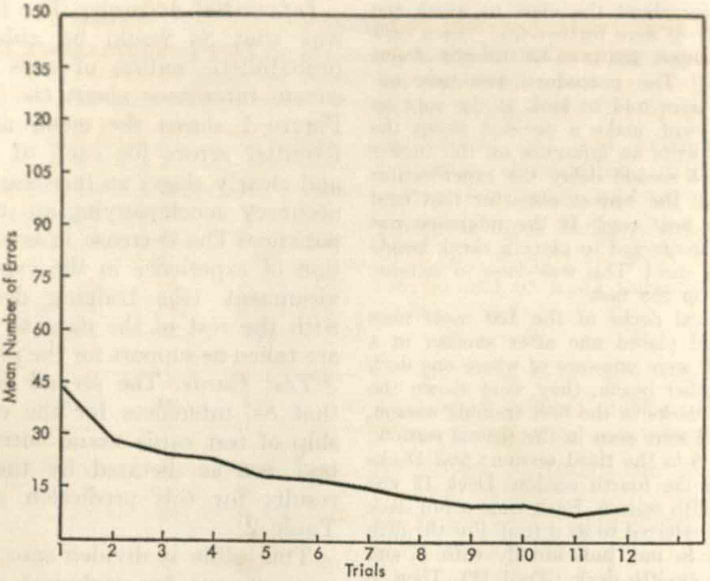


FIG. 1. Mean number of errors made on each trial by 10 *Ss*: Experiment I.

that *Ss* inferences for test cards correspond to the best bets yielded by the model.

Confirming and Infirming Cards. The third prediction was that *Ss'* inferential accuracy for confirming cases should increase with increased experience with the experimental environment and the fourth was that there would be a decrease in inferential accuracy for infirming cases.

In order to evaluate the accuracy of these two predictions the proportion of *Ss* giving each of the three response alternatives was computed for each card for each trial.[5] The cards were then divided into confirming and infirming cards, and the mean proportion of *Ss* giving each response alternative was found for the two kinds of cards. These are plotted in Figures 2, 3, and 4.

In Figure 2 the mean proportion of *Ss* giving the correct class for the confirming cards is compared with the mean proportion of *Ss* giving the correct class for infirming cards on each of the 12 trials. These results support the third prediction, inferential accuracy for confirming cards clearly increases as *Ss'* experience with the experimental environment increases. The fourth experimental prediction is contradicted by the results in Figure 2; the inferential accuracy for infirming cards *increases* with experience in the experimental environment.

The fifth prediction was that errors on the infirming cards would result from assigning those cards to the class which the model dictates as the best bet, rather than to the correct class. The results for this prediction are given by a comparison of Figures 3 and 4. For the confirming cards the curves in Figure 3 represent the mean proportion of *Ss* giving the correct response (which is also the theoretical best bet) and the other two responses for each trial. The top curve indicates the correct response, and the bottom two curves indicate the two incorrect alternatives. For the infirming cards the curve in Figure 4 which starts in the middle and ends high is the mean proportion of *Ss* who gave the correct response (which is not the theoretical best bet) on each of the 12 trials. The curve which starts high and ends in the middle is the mean proportion of *Ss* who gave as their inference the alternative which, while incorrect, was the theoretical best bet. The curve that starts low and ends lower is the mean proportion of *S*s who on each trial gave as their inference for the infirming cards the alternative which was neither correct nor the theoretical best bet. The fifth prediction is supported by these data; for the most part, errors for infirming cards resulted from assigning the cards to the class which was the theoretical best bet rather than to the correct class.

---

[5] The data analysis dealt only with the proportions yielded for the first time a card appeared on a trial. For example, there are six d37 cards in the deck; the analysis utilizes the results for the first d37 to appear on each trial.

Discussion

These results indicate that *S*s learned to utilize probabilistic cues to make accurate inferences for *O*s (see Figure 1). Moreover, by using the model it was possible to make fairly accurate predictions of *S*s' responses for test *O*s which they had never seen before (see Table 2). For those *O*s which the model designated to be confirming cases *S*s quickly came to give the correct class, and the alternative inferences dropped out. However, the results for the infirming cases were not so clear-cut. For these cases *S*s tended to give the theoretical best bet in the early trials, but later they tended to give the correct—low $E(k)$ value—response. This result contradicts the hypothesis that inferential accuracy would decrease for infirming cases. However, the infirming card results support the fifth hypothesis; the incorrect responses for infirming cases are usually the theoretical best bet as determined by the model.

While the results of the experiment indicated that the theoretical approach outlined in the introduction is a promising way of viewing inference behavior, the small number of classes (three) and the small number of cue values (nine) on each cue dimension place restraint upon acceptance of the results. It seemed possible that these results might not be obtained a second time if more classes and longer cue dimensions were used. And, of course, the major question is how the fourth prediction, that there would be a decrease in inferential accuracy over trials for infirming cards, would fare in a replication of the experiment.

Furthermore, the small number of *S*s and the way in which the test cards were

TABLE 2

TEST CARD INFERENCE RESULTS FOR EXPERIMENT I

Test card	Class	$E(k)$	Response distribution	χ^2
g66	**R**	**.45**	3	3.80 ($n = 10$)
	Y	.27	6	
	B	.28	1	
f35	R	.30	4	.79 ($n = 10$)
	Y	**.49**	4	
	B	.21	2	
c27	R	.02	0	20.02 ($n = 10$)***
	Y	**.64**	10	
	B	.33	0	
d54	**R**	**.39**	5	5.01 ($n = 10$)
	Y	.26	0	
	B	.35	5	
f5-	**R**	**.58**	7	7.40 ($n = 10$)*
	Y	.39	3	
	B	.02	0	
c-6	R	.06	0	5.60 ($n = 10$)
	Y	.34	4	
	B	**.60**	6	
-65	R	.36	1	7.30 ($n = 9$)*
	Y	.16	1	
	B	**.49**	7	
d2-	R	.04	2	10.40 ($n = 10$)**
	Y	**.75**	8	
	B	.22	0	
1-2	**R**	**.97**	10	20.02 ($n = 10$)***
	Y	.00	0	
	B	.03	0	
e2-	R	.14	2	10.40 ($n = 10$)**
	Y	**.82**	8	
	B	.05	0	
g-4	**R**	**.57**	10	20.02 ($n = 10$)***
	Y	.14	0	
	B	.28	0	
-28	R	.04	0	20.02 ($n = 10$)***
	Y	**.93**	10	
	B	.04	0	
h-1	**R**	**.97**	10	20.02 ($n = 10$)***
	Y	.03	0	
	B	.00	0	
a9-	R	.00	0	20.02 ($n = 10$)***
	Y	.00	0	
	B	**1.00**	10	

* $p < .05$.
** $p < .02$.
*** $p < .001$.

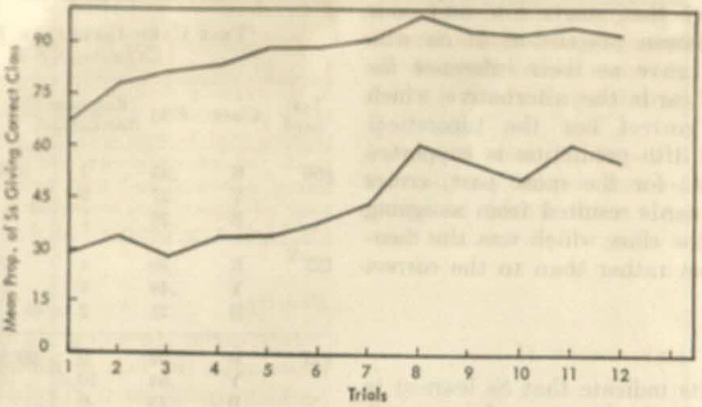


FIG. 2. Mean proportion of *Ss* giving correct class as their responses: Experiment I. (Upper curve = confirming cards. Lower curve = infirming cards.)

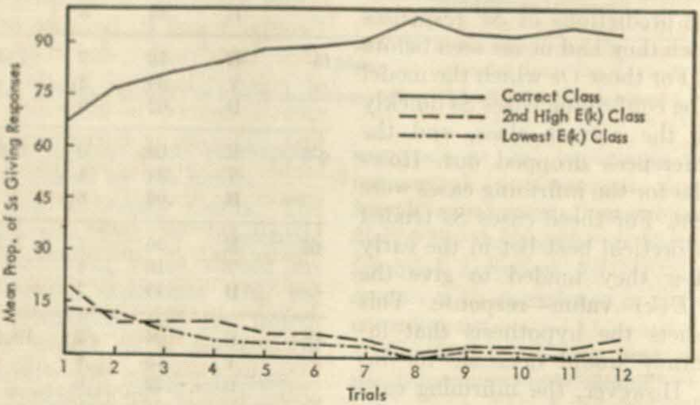


FIG. 3. Confirming card inference results for Experiment I.

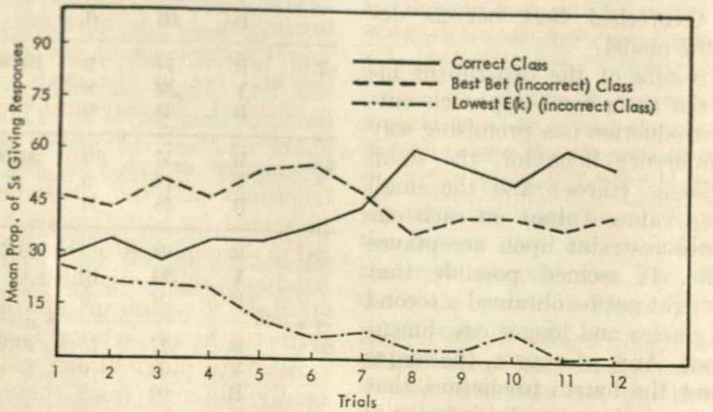


FIG. 4. Infirming card inference results for Experiment I.

presented to *Ss* also place limitations upon the acceptance of the results of the study. In this experiment *Ss* were shown the test cards in a separate session following the learning trials; for some *Ss* this session was almost 48 hours after the final training trial. This delay may have influenced the results on the test cards. An alternative to this procedure would be to place the test cards in the training deck on the last trial so that *Ss* would make inferences about them in the course of making inferences about the deck cards.

Because of these limitations in experimental procedure and because a replication of the obtained results would be strong support for the utility of the present theoretical approach, another experiment was conducted.

EXPERIMENT II

Construction of the Experimental Environment

The research strategy used in the first experiment also guided the procedure used in this one. The deck was larger in that it consisted of 200, rather than 120 cards. Each card belonged to one of five different classes (red, yellow, blue, green, or orange) and there were 40 cards in each class. The cue dimensions were expanded to 12 values: *a* through *l* on the letter dimension, 1 through 12 on the number dimension, and Positions 1 through 12 (in a clockwise direction) on the pointer dimension. The method used in the first experiment to select the cue combination for each card and the number of each kind of card in the deck was also used for this experiment. (Contingency coefficients for the relationship between each cue dimension and the classes were .76 for letter, .70 for number, and .75 for pointer, indicating that no one of the dimensions was a better predictor of the cards' classes than the other dimensions.) The matrix marginal distributions for each class and the values of $P(k/d)$ for this deck of cards are presented in Table 3. This deck was not an extension of the one used in the first experiment. The method of construction is the same, but the deck and the cue distributions were completely new.

Application of the model to the deck of cards yielded 78 different kinds of cards which were confirming cards and 14 which were infirming cards. In addition to the deck cards 15 test cards were constructed, 12 of which had three cue values and 3 of which were blank.

The Ss

The *Ss* were United States Naval Aviation Cadets of approximately the same age and intelligence as *Ss* used in the first experiment. Because part of the group of *Ss* transferred before the experiment was completed, the data are reported for two groups. Group A had 13 training sessions, and Group B had only 8 sessions. The *Ss* using all three cue dimensions were selected in the same way as in Experiment I. This resulted in the use of data from 16 out of 39 *Ss* in Group A and 20 out of 41 *Ss* in Group B.

Procedure

The procedure was the same as for Experiment I except that the test cards were placed randomly in the training deck on the last trial rather than being presented at a special session. For this last training trial *Ss* were told that they would not be given the correct answer for the cards. The *E* merely presented each card, paused for *Ss* to make their inferences, then presented the next card. The test cards were presented as they came up in the deck.

Results

Inferential Accuracy. As in the previous experiment the first prediction was that *Ss* could learn to use the probabilistic cues to make accurate inferences. The curves in Figure 5 represent the mean number of errors made by *Ss* on each trial, for both Groups A and B. Because of the expanded deck (200 cards) and the greater number of classes, learning in the second experiment appears to be more difficult than in the first. The error curves in Figure 5 start higher and end higher than the comparable curves for the first experiment (see Figure 1). In general, however, these results indicate that *Ss'* inferences become increasingly accurate as training progresses. This decrease in errors with training is essentially the same as the corresponding results of the first experiment and, together with the other results reported here, supports the first prediction.

Test Cards. The results for the second prediction, that *Ss'* inferences for test cards would correspond to the class designated the best bet by the model, are presented in Table 4. The arrangement of the table is similar to the one presented for the first experiment except that the response distributions and the chi-square results are presented for both Groups A and B.

For Group A 9 of the 12 response distributions differ significantly from chance. Of these 9 significant distributions 6 conform to the prediction. This leaves 3 test cards with distributions which are significantly different from chance, but do not conform to the predictions based on the model. In short, *Ss* selected the theoretical best bet on only 6 of the 12 test cards.

For Group B 9 of the 12 response distributions are significantly different from chance. Of these 9 significant distributions

TABLE 3

FREQUENCY DISTRIBUTIONS OF CUE VALUES' ASSOCIATION WITH EACH CLASS AND THE CORRESPONDING VALUES OF $P(k/d)$: EXPERIMENT II

Class	Letter											
	A	B	C	D	E	F	G	H	I	J	K	L
Red												
f	0	0	0	0	0	0	1	3	6	10	14	6
$P(k/d)$	.00	.00	.00	.00	.00	.00	.04	.19	.40	.77	.78	1.00
Yellow												
f	0	1	4	8	14	8	4	1	0	0	0	0
$P(k/d)$	.00	.07	.29	.57	.54	.27	.16	.06	.00	.00	.00	.00
Blue												
f	8	14	10	6	2	0	0	0	0	0	0	0
$P(k/d)$	1.00	.93	.71	.43	.08	.00	.00	.00	.00	.00	.00	.00
Green												
f	0	0	0	0	6	14	8	6	6	0	0	0
$P(k/d)$	.00	.00	.00	.00	.23	.47	.32	.38	.40	.00	.00	.00
Orange												
f	0	0	0	0	4	8	12	6	3	3	4	0
$P(k/d)$	.00	.00	.00	.00	.15	.27	.48	.38	.20	.23	.22	.00
	Number											
	1	2	3	4	5	6	7	8	9	10	11	12
Red												
f	0	1	4	8	14	8	4	1	0	0	0	0
$P(k/d)$	.00	.04	.14	.31	.53	.36	.29	.14	.00	.00	.00	.00
Yellow												
f	4	14	10	6	4	2	0	0	0	0	0	0
$P(k/d)$	.40	.61	.36	.23	.15	.09	.00	.00	.00	.00	.00	.00
Blue												
f	0	0	0	0	0	0	0	1	5	9	11	14
$P(k/d)$	.00	.00	.00	.00	.00	.00	.00	.14	.50	1.00	1.00	1.00
Green												
f	0	0	0	0	8	12	10	5	5	0	0	0
$P(k/d)$	.00	.00	.00	.00	.31	.54	.71	.71	.50	.00	.00	.00
Orange												
f	6	8	14	12	0	0	0	0	0	0	0	0
$P(k/d)$	.60	.34	.50	.46	.00	.00	.00	.00	.00	.00	.00	.00
	Pointer											
	1	2	3	4	5	6	7	8	9	10	11	12
Red												
f	4	14	10	6	4	2	0	0	0	0	0	0
$P(k/d)$	1.00	.56	.36	.30	.17	.10	.00	.00	.00	.00	.00	.00
Yellow												
f	0	0	0	0	0	6	10	14	6	4	0	0
$P(k/d)$	.00	.00	.00	.00	.00	.28	.59	.78	.50	.22	.00	.00
Blue												
f	0	1	4	8	14	8	4	1	0	0	0	0
$P(k/d)$	.00	.04	.14	.40	.61	.38	.24	.06	.00	.00	.00	.00
Green												
f	0	10	14	6	5	5	0	0	0	0	0	0
$P(k/d)$	.00	.40	.50	.30	.22	.24	.00	.00	.00	.00	.00	.00
Orange												
f	0	0	0	0	0	0	3	3	6	14	10	4
$P(k/d)$	.00	.00	.00	.00	.00	.00	.18	.17	.50	.78	1.00	1.00

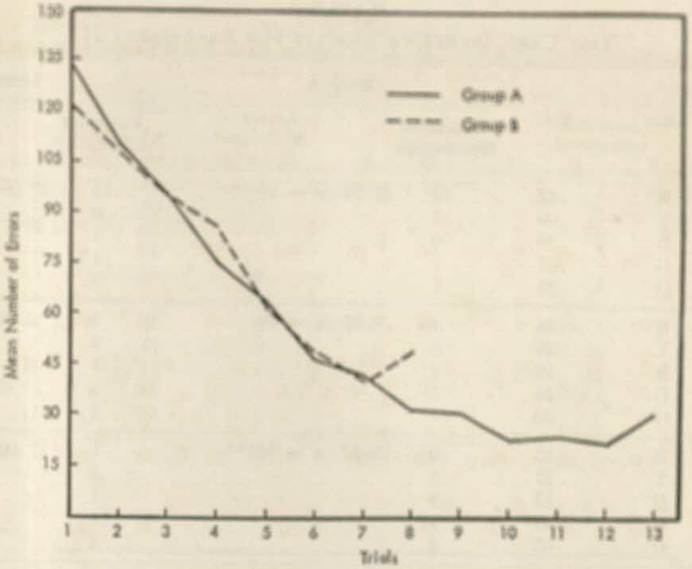


Fig. 5. Mean number of errors made on each trial by 16 *Ss* in Group A and 20 *Ss* in Group B: Experiment II.

7 conform to the prediction. This leaves 2 test cards with distributions which are significantly different from chance, but which do not conform to the predictions based on the model. Thus, *Ss* in Group B selected the model designated best bet on only 7 of the 12 test cards. At best these results only partially confirm the previous results and are only equivocal support for the prediction that test card inferences correspond to the theoretical best bet.

Confirming and Infirming Cards. The learning curves for the confirming and for the infirming cards, for both groups of *Ss*, are compared in Figure 6. The difference between the curves for the two kinds of cards is similar to the corresponding results for the first experiment and supports the hypothesis that the probabilistic aspect of the cues are a factor influencing the accuracy of *Ss*' inferences.

Figures 7 and 8 contain the results for the confirming cards for Groups A and B, respectively. In these graphs the top curve represents the class—correct—with the highest $E(k)$ value, the second curve the second highest $E(k)$—incorrect—class, the third curve the third highest $E(k)$—incorrect—class, etc. (The slight drop in accuracy at the end of training occurs during the last trial on which *Ss* were not told the correct answers for the cards.) Again, the results are highly similar to the results from the first experiment and support the prediction that *Ss*' inferential accuracy for confirming cards increases as experience in the experimental environment increases.

Figures 9 and 10 contain the results for the infirming cards for the two groups, respectively. Here the curve which is highest on the last trials is for the correct class; this is not, however, the highest $E(k)$ class according to the model. The second curve is for the incorrect class, which the model indicates is the best bet. The third, fourth, and fifth curves are the third highest, fourth highest, and fifth highest $E(k)$—incorrect—classes. These results, which are similar to the results from the first experiment, prove the fourth prediction to be wrong. The results for all three groups of *Ss* (the students and both groups of cadets) show that inferential accuracy did *not* decrease for infirming cards.

The fifth prediction, that *Ss*' errors on infirming cards would result from giving the best bet rather than the correct answer, is supported in Figures 9 and 10. Here too

TABLE 4

TEST CARD INFERENCE RESULTS FOR EXPERIMENT II

Test card	Class	$E(k)$	Group A		Group B	
			Response distribution	χ^2	Response distribution	χ^2
K45	**R**	.42	13	37.74 (n = 16)***	15	39.00 (n = 20)***
	Y	.08	1		0	
	B	.20	0		1	
	G	.07	1		1	
	O	.23	1		3	
D52	**R**	.36	4	8.37 (n = 16)	5	2.50 (n = 20)
	Y	.24	3		6	
	B	.16	2		3	
	G	.24	7		4	
	O	.00	0		2	
E76	R	.13	0	18.37 (n = 16)**	1	17.00 (n = 20)**
	Y	.27	5		3	
	B	.15	0		1	
	G	.39	9		11	
	O	.05	2		4	
G25	R	.08	6	8.37 (n = 16)	3	8.50 (n = 20)
	Y	.26	5		2	
	B	.20	0		1	
	G	.18	4		8	
	O	.27	1		6	
B47	R	.10	1	27.12 (n = 16)**	1	28.50 (n = 20)***
	Y	.30	11		12	
	B	.39	4		7	
	G	.00	0		0	
	O	.21	0		0	
F23	R	.13	6	8.37 (n = 16)	3	6.50 (n = 20)
	Y	.29	4		5	
	B	.05	0		2	
	G	.32	5		8	
	O	.20	1		2	
H58	R	.24	0	17.12 (n = 16)**	1	18.50 (n = 20)***
	Y	.33	4		10	
	B	.03	0		0	
	G	.23	3		2	
	O	.18	9		7	
K410	R	.36	0	54.62 (n = 16)***	2	32.50 (n = 20)***
	Y	.15	1		3	
	B	.00	0		0	
	G	.00	0		1	
	O	.49	15		14	
D54	**R**	.28	2	30.87 (n = 16)***	2	11.50 (n = 20)*
	Y	.24	0		5	
	B	.28	1		4	
	G	.20	12		9	
	O	.00	1		0	
I43	**R**	.38	14	46.50 (n = 16)***	17	53.00 (n = 20)***
	Y	.08	0		1	
	B	.05	0		0	
	G	.30	2		1	
	O	.22	0		1	

Table 4—Continued

Test card	Class	$E(k)$	Group A Response distribution	Group A χ^2	Group B Response distribution	Group B χ^2
F47	R	.10	1	45.87 ($n = 16$)***	1	38.50 ($n = 20$)***
	Y	.38	14		15	
	B	.08	0		0	
	G	.16	1		2	
	O	.30	0		2	
E16	R	.03	6	15.87 ($n = 16$)**	1	13.50 ($n = 20$)**
	Y	.41	8		10	
	B	.15	1		1	
	G	.16	0		4	
	O	.25	1		4	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

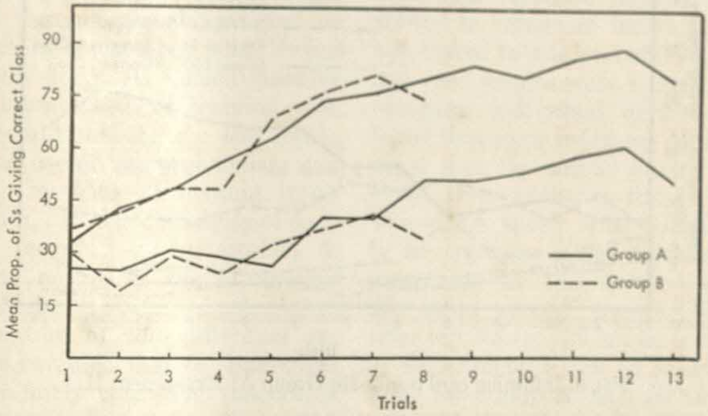


FIG. 6. Mean proportion of *S*s giving the correct class as their response: Experiment II. (Upper curve = confirming cards. Lower curve = infirming cards.)

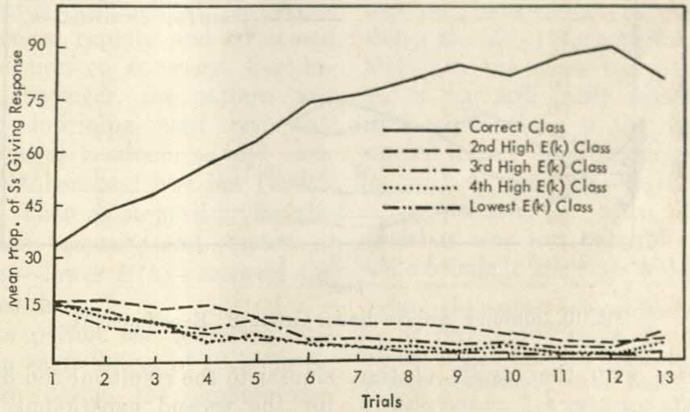


FIG. 7. Confirming card results for Group A: Experiment II.

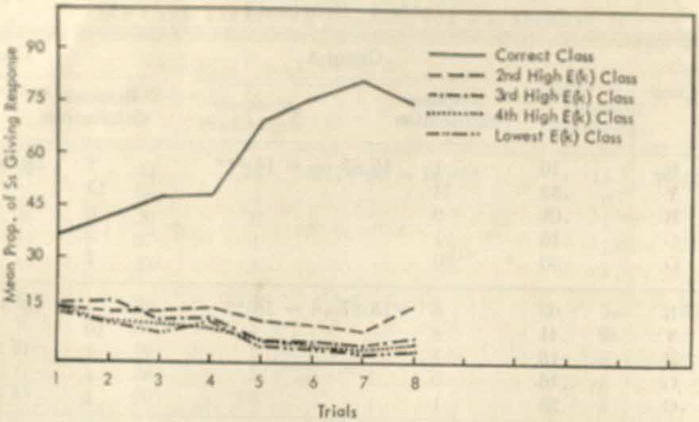


Fig. 8. Confirming card results for Group B: Experiment II.

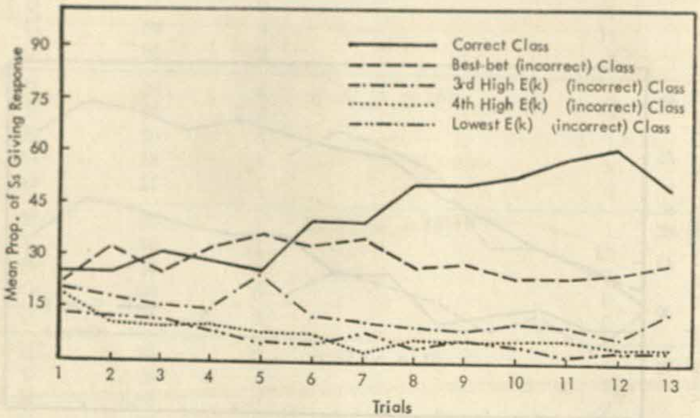


Fig. 9. Infirming card results for Group A: Experiment II.

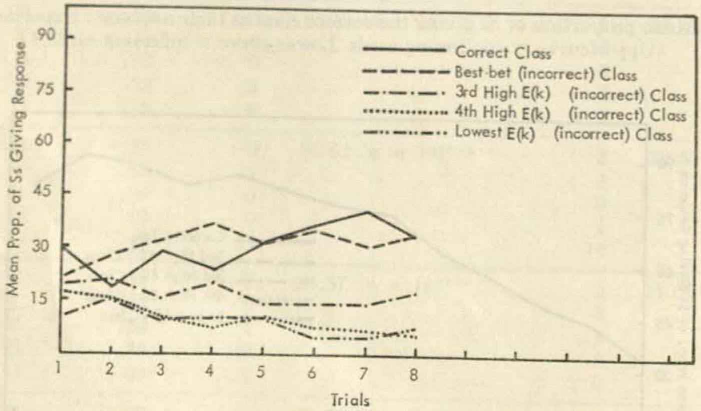


Fig. 10. Infirming card results for Group B: Experiment II.

the results agree with the results of the first experiment.

Alternative Inference Methods. The results of the second experiment are highly similar to the results of the first except that for the second experiment the test card responses did not conform as closely to the model as they did in the first experiment.

This lack of predictability for test cards, along with the consistent failure of the data to support Prediction 4, indicates that the model does not fully account for *Ss'* inferences.

While the results show that the model is not wholly correct, they also show that it is not wholly wrong. First, the theoretical best bet for infirming cards was most frequently chosen in the early stages of training (see Figures 4, 9, and 10), even though the correct answer became more frequent as training progressed. Second, Predictions 2 and 4, that *Ss'* test card inferences would correspond to the model-derived best bet and that accuracy would decrease for infirming cards, were only partially incorrect. Some of the test card results conformed to the theoretical best bet. These results indicate that the cue probabilism had an influence upon inferences but that it was not the whole story. It seemed possible that in the later stages of training and, under some conditions, for the test cards, *Ss* relinquished use of cue probabilism and adopted other methods of making inferences. Therefore, a detailed analysis of each *S's* inference records was undertaken to investigate the nature of the alternative inference methods.

The examination of the individual response records revealed that *Ss'* inferences all followed a fairly consistent pattern, a pattern which is reflected in the curves presented earlier (see Figures 2–4 and 6–10). In the beginning of training the responses were essentially random. Then, for confirming cards, inferential accuracy started to increase rapidly and continued to rise toward perfect accuracy. For infirming cards, however, the pattern was different. The infirming card responses started with initial randomness and then changed to the theoretical best bet (which was incorrect). Then *Ss* stopped giving the best bet as their response and started to give the correct—lower $E(k)$—answers for the infirming cards.

These results permit the following hypothesis. In the early stages of training *Ss* began to learn about the probabilistic aspects of the cues (until these were learned the responses were random), then, as their distribution hypotheses developed *Ss* began to give the best bet as their inferences. However, as training progressed and the deck cards were seen repeatedly, *Ss* became familiar with the individual cards and their component cues. As this happened, *Ss* could increasingly rely upon *recognition* of individual cards and memory of their correct (not the best bet) classes. This change in inference method was evidenced in the change from best bet to correct answers for the infirming cases but, of course, it was not reflected for confirming cards because for these cards the correct and best-bet answers are the same.[6]

In addition to clarifying the response pattern for the infirming cards the analysis of individual response records revealed how ability to recognize individual deck cards was linked to cue probabilism and how the test card results were related to ability to recognize individual deck cards. It was found that some infirming cards were recognized (i.e., the correct answer was consistently given) earlier in training than others. The mean speed with which cards came to be recognized was correlated with the magnitude of the $E(k)$ associated with the cards' *correct* classes. These correlations are .93 ($n = 9$ cards, $p < .001$) for *Ss* in Experiment I, .79 ($n = 14$ cards, $p < .001$) for Group A, and .52 ($n = 14$ cards, $p \cong .07$) for Group B (Spearman rank-order correlation, two-tailed probabilities). These results indicate that *Ss'* use of recognition as a method for making inferences was not independent of cue probabilism. When the $E(k)$ for the correct class was high, i.e., the class was not the best bet but it was still fairly reasonable that the card could belong to the class, *Ss* quickly started to rely upon recognition of the card in order to determine its class. When the $E(k)$ was low, i.e., when in terms of the model it was less probable that the card could belong to the class which was actually correct, *Ss* waited until later in training to utilize recognition.[7]

---

[6] An unsuccessful attempt was made to account for *Ss'* inferences in terms of a progression from the use of cues as single elements, through the use of pairs as elements, to the use of all three cues as a single pattern. The existence of the second stage, use of pairs of cues, was not supported by the data.

The impression should not be given that *Ss* used the model or recognition in an all-or-none way; indeed, there were great individual differences in the degree to which recognition was utilized. The degree to which it was used can be estimated by the relative numbers of correct versus best-bet answers given for the infirming cards. Some *Ss* appeared to start to use recognition early in training; others used it only for those infirming cards for which the $E(k)$ for the correct class was high. The nonrecognizing *Ss*, those who depended more heavily upon cue probabilism, tended to give the best bets as their responses to infirming cards all through training even though they were frequently wrong. These individual differences in the use of recognition provided the key to understanding the test card results.

The test card results show that for some cards there were frequently two answers which were more commonly given. One of these was usually the theoretical best bet; the other was frequently the answer which was correct for a deck card which possessed cue values similar to the test card's cues. This suggested that *Ss* who depended upon recognition in the training trials might have tended to base their test card answers on the similarity between the test card and a familiar card from the training deck. To investigate this, the proportion of correct responses for infirming cards was used as an indication of the degree to which each *S* depended upon recognition during training. This was correlated with the proportion of *S*'s test card inferences which could be accounted for by the similarity between the test card and a card from the deck.[8] These correlations were .71 ($n =$ 10 *Ss*, $p < .05$) for *Ss* in Experiment I, .37 ($n = 16$ *Ss*, $p \cong .15$) for Group A, and .51 ($n = 20$ *Ss*, $p < .05$) for Group B (Spearman rank-order correlation, two-tailed probabilities). These correlations support the hypothesis that *Ss* who rely on recognition during training also tend to make "similarity-based" inferences for the test cards rather than "model-based" inferences. It is also important to note that the test cards did not have to be similar to infirming cards to receive similarity-based responses; similarity to confirming cards also influenced *Ss*' inferences. This implies that the familiarity provided by the training trials permitted these *Ss* to remember both the confirming and the infirming cards and to utilize this information for their test card inferences.

Probability Matching and Theory. The paradigm for probability learning studies is that *Ss* are given a ready signal and asked to predict which of two or more alternative events will occur. The ready signal may be a light, a buzzer, a verbal "ready," or the like; and the events may be different lights on a display, symbols on cards, digits writen on a blackboard, etc. The events occur randomly and with different relative frequencies.

It can be argued that the present theory is applicable to the probability learning situation; that the ready signal is a one-valued cue, $d = 1$, and that the subsequent events are classes, $k = a$ through i. In this situation the theory would predict that *Ss* should tend to maximize, i.e., to always select the most frequently occurring event. This, of course, is not what usually happens. In most of the probability learning studies *Ss*' frequency of response approximates the relative frequency of occurrence of the corresponding events.

It is possible that differences between the experimental situations in the present research and in the probability learning studies make the theory inapplicable to the latter. However, it is possible that the distributions of responses for cards in the present experiment result from *Ss*' tendency to match their response frequencies to the theoretical probabilities. Therefore, the following analysis was performed.

---

[7] The finding that the speed of adopting recognition is related to cue probabilism is further supported by the fact that the frequency of the occurrence of a given kind of card in the deck (e.g., there were six E52 cards, three F73 cards, etc.) does not correlate with the speed of adopting recognition for either confirming or infirming cards.

[8] To be considered similar to a test card a deck card could have only one of the three cue values different from the cues on the test card. There were 5 of the 14 Experiment I test cards and 6 of the 12 Experiment II test cards for which the model-based best bet and the answer dictated by similarity to a deck card were not the same answer.

If *Ss* tend to maximize rather than to match probabilities, the proportion of *Ss* selecting the best bets for the test cards should be significantly higher than the best bets' probabilities. Using those test cards which had nonchance response distributions, this difference was significant at or below the .05 level for all three groups of *Ss* (Mann-Whitney *U* test, one-tailed).

By the same reasoning, if *Ss* are matching probabilities rather than maximizing on the training cards, the response curves for the best bet (see Figures 2-4 and 6-10) should approximate the average of the best-bet probabilities for the corresponding cards. For Experiment I these averages are .69 for the best bet for the confirming cards, .50 for the best bet for the infirming cards, and .26 for the correct class for the infirming cards. For Experiment II these averages are .54 for the best bet for the confirming cards, .39 for the best bet, and .22 for the correct class of infirming cards. Comparison of the curves in the figures with these values provides no support for the probability matching hypothesis. Thus, for both test cards and for the training cards there is no evidence of *Ss*' responses matching the theoretical probabilities.

Discussion

In general, the present theory appears to provide a promising line of inquiry. The results indicate that for unfamiliar objects cue probabilism is an important determinant of *Ss*' inferences and that with increased familiarity the use of recognition and of similarity becomes important. This conclusion is supported, first, by the curves representing the infirming cards. Figures 4, 9, and 10 show that early in training the theoretical best bet was always given by the highest proportion of *Ss*. Later, as training proceeded, and *Ss* supposedly became able to recognize the cards, the response curve for the correct class rose and exceeded the curve for the best bet.

The test card results also support the conclusion. The response distributions for the test cards (see Tables 2 and 4) frequently have two high points: one corresponding to the theoretical best bet, and the other corresponding to the "similarity-based" response. Unless they are the similarity-based responses, the classes having the lowest theoretical $E(k)$ are infrequently chosen.

The influence of cue probabilism on *Ss*' utilization of recognition is demonstrated by *Ss*' individual response records. These records reveal that the magnitudes of the $E(k)$ values derived from the model correspond to the order in which *Ss* adopt recognition for classifying infirming cards. The individual response records also showed that *Ss* who depended most heavily upon recognition during training also depended most heavily upon similarity in responding to the test cards. The latter finding indicates that there are individual differences in how *Ss* utilize the various alternative inference methods; differences which are perhaps linked with motivation, intelligence, and other such factors.

The finding that the present results do not indicate that *Ss* were matching probabilities may be due to some fundamental difference between the present experimental situation and the usual probability learning situation; a difference which causes *Ss* to use different strategies, decision rules, sets, or the like, in the two situations. When discussing the differences between the two situations it is well to keep in mind that neither situation is as simple as it first appears. A number of studies (e.g., Cotton & Rechtschaffen, 1958; Gardner, 1958; Hake & Hyman, 1953; Jarvik, 1951; Jones 1961; McCormack, 1959) have demonstrated various factors which contribute to the underlying complexity of the probability learning situation. No doubt similar analyses would prove the present situation to be equally complex. One important aspect of the probability learning situation is *Ss*' tendency to ignore the independent existence of each trial (Hake & Hyman, 1953; Jarvik, 1951). Instead the experiment apparently is looked upon as a unit in itself of which one aspect, the sequence of outcomes, is variable. The variable outcomes are seldom assumed to be random, however, and *Ss*' are particularly alert for real or imagined patterns in the sequence. Attention to patterns is not surprising because *Ss* seldom encounter purely random sequences of highly similar, cue-impoverished events in the normal environment. When such a sequence does occur, as in gambling or perhaps in the patterns of lights which flicker on and off on warning signals, advertising signs, or strings of flickering Christmas tree lights, there is a very strong tendency to impose order upon the events. The imposed order is probably not independent of whatever parameters determine the occurrence of the events, but it is keyed, nonetheless, to the sequence rather than to the individual events as distinct *O*s.

In the present experiment an effort was made to prevent *Ss* from relying upon the order of *Os*' occurrence in the deck. The *Ss* were instructed to think of each card as an individual *O* (cf. page 7), and subsequent comments made by various *Ss* indicated that for the most part they followed the instructions. These instructions may account for the apparent lack of probability matching in the present results.

REFERENCES

ATTNEAVE, F. Psychological probability as a function of experience frequency. *J. exp. Psychol.*, 1953, **46**, 81–86.

BEACH, L. R., & WERTHEIMER, M. A free response approach to the study of person cognition. *J. abnorm. soc. Psychol.*, 1961, **62**, 367–374.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.

BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.

BRUNSWIK, E. *Perception and the representative design of experiments.* Berkeley: Univer. California Press, 1956.

COTTON, J. W., & RECHTSCHAFFEN, A. Replication report: Two- and three-choice verbal conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

EDWARDS, W. Behavioral decision theory. *Annu. Rev. Psychol.*, 1961, **12**, 473–497.

GARDNER, R. A. Multiple-choice decision behavior. *Amer. J. Psychol.*, 1958, **71**, 710–717.

GOODNOW, R. E. Utilization of partially valid cues in perceptual identification. Unpublished doctoral dissertation, Harvard University, 1954.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

HAYS, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri, & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958.

JARVIK, M. E. Probability learning and a negative recency effect on the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

JONES, JOAN E. Deviations from matching behavior in probability learning. *J. exp. Psychol.*, 1961, **52**, 335–345.

MCCORMACK, P. D. Spatial generalization and probability-learning in a five-choice situation. *Amer. J. Psychol.*, 1959, **72**, 135–138.

MEEHL, P. E. *Clinical vs statistical prediction: A theoretical analysis and review of the evidence.* Minneapolis: Univer. Minnesota Press, 1954.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PRESTON, M. G., & BARATTA, P. An experimental study of the auction-value of an uncertain outcome. *Amer. J. Psychol.*, 1948, **61**, 183–193.

RAPPOPORT, L. Multiple probability learning and personality. Unpublished master's thesis, University of Colorado, 1961.

SARBIN, T. R. A contribution to the study of actuarial and individual methods of prediction. *Amer. J. Sociol.*, 1943, **48**, 593–602.

SARBIN, T. R., TAFT, R., & BAILY, D. E. *Clinical inference and cognitive theory.* New York: Holt, 1960.

SIGNELL, K. A. Inferential behavior in person-perception: Cognitive structure, information-seeking strategy and differential dependence on cognitive dimensions. Unpublished master's thesis, University of Colorado, 1961.

SMEDSLUND, J. *Multiple-probability learning: An inquiry into the origins of perception.* Oslo: Akademisk, 1955.

SUMMERS, S. A. Learning of responses to multiple weighted cues. *J. exp. Psychol.*, 1962, **64**, 29–34.

WERTHEIMER, M. Values in person cognition. In D Willner (Ed.), *Decisions, values, and groups.* Vol. I. New York: Pergamon Press, 1960.

(Received March 7, 1963)

Vol. 78, No. 6 Whole No. 583, 1964

Psychological Monographs: General and Applied

RECOGNITION, ASSIMILATION, AND IDENTIFICATION OF OBJECTS[1]

LEE ROY BEACH

Office of Naval Research, Washington, D. C.

The present theory, based on the results of the author's previously reported research, concerns the ways in which Ss make inferences about what objects (Os) are, i.e., about their class memberships, and about their covert properties, i.e., about their unknown cue values. Mechanisms and conditions are described for 3 inference methods: for familiar Os (recognition), for unfamiliar Os which are similar to a familiar O (assimilation), and for unfamiliar Os (identification). In addition, the way in which this knowledge is retained and how it is revised in light of new information are also discussed.

THE course of a man's life is intricately interwoven with the multitude of persons, things, and events which he encounters, some repeatedly and some only once. Interaction with these objects (*O*s) presumes knowledge about them, knowledge which allows the subject (*S*) to plan his part of the interaction properly and to have some notion of how *O* will react to his actions. Knowledge about *O*s can be divided into two general categories: (*a*) knowledge about what kind of *O* it is, i.e., the general class of *O*s to which it belongs; and (*b*) knowledge about the nature of *O*, i.e., *O*'s properties or characteristics. When an *O* is first encountered *S* obtains some information about it through immediately apparent characteristics, such as contour, color, and sound, from which he attempts to determine the kind of *O* it is. Then the knowledge about what *O* is, together with the knowledge about its immediately apparent characteristics, can be used to decide about *O*'s covert properties. The present theory is concerned with how *S*s utilize information about the known attributes of *O*s in order to make inferences about *O*s' unknown attributes.

In brief, the theory assumes that each *S* possesses a multidimensional cognitive space which is defined by the cue dimensions to which he attends. All of the cue values *S* knows to be possessed by *O* define a location in this space, the cognitive representation of *O*, called θ. If, when an *O* is referred to the space, a θ already exists at or extremely near the location, *O* is said to be *recognized* as the previously experienced *O* corresponding to θ. Then the cue values and the class associated with θ can be assumed to be applicable to the recognized *O*. If the location defined by *O*'s cue values falls somewhat further from a θ in the space, *S* can still assume θ and *O* to be highly similar in class membership and in covert cue values. Here the class and covert cue values associated with the similar θ can be *assimilated* for *O* and assumed to be sufficient until further information indicates the contrary. If the location defined by *O*'s cue values does not fall close to a θ in the space, *S* can fall back on a third method called *identification*, which relies on the statistical properties of groups of θs which possess cue values similar to the ones *O* is known to possess. These three inference processes comprise the major portion of the theory. The remainder consists of the methods by which the inferences are checked against subsequent information and by which revisions are made in an effort to maintain accuracy.

---

[1] Opinions or conclusions contained in this report are those of the author and not to be construed as necessarily reflecting the views or the endorsement of the Navy Department. The valuable help and advice of K. R. Hammond, M. Wertheimer, M. B. Jones, J. R. Berkshire, D. E. Bailey, Alice Garman and Elizabeth McPeters is gratefully acknowledged. This work was done at the United States Naval School of Aviation Medicine, Pensacola, Florida.

ATTRIBUTES OF *O*s

In their *Study of Thinking* Bruner, Goodnow, and Austin (1956) define an *attribute* of an *O* as

> any discriminable feature of an [*O*] that is susceptible of some discriminable variation from [*O*] to [*O*] [p. 26].

These authors also quote Boring's (1942) similar definition:

> A stone is shape, color, weight, and kind of substance in complicated relation. When such descriptive ultimates are general properties which can vary continuously or discretely, when they are, in short, parameters, they may, if one chooses, be called attributes of the object described [p. 25].

As these definitions indicate, *O*s are defined by their attributes, and attributes are those aspects of *O*s which permit discriminations among them. It is assumed that there are two types of attributes associated with *O*s, classes and cue values.

Classes

A *class* is an attribute of *O*s which permits only dichotomous (presence or absence) judgments about its association with any one *O*. It is a nominal attribute of an *O* which does not permit discriminations among *O*s which possess it. For example, knowing that two *O*s are members of the class "cow" does not aid in discriminating between them—discriminations are based upon other attributes such as color, size, and markings, which are cues. Classes exist as a result of the psychological operation of putting *O*s together into various groups and of assigning these groups specific names. While they can be subdivided on the basis of discriminations made on cue dimensions, e.g., red houses and white houses, the class still remains the group of *O*s to which the discrimination refers, and at no time can various members of one class be discriminated from one another on the basis of knowledge of the class alone.

Class Hierarchy

It is obvious that an *O* is associated with many different classes and that knowledge of its membership in one class often implies membership in a number of other classes. This is because classes exist in a hierarchy, which runs from an infinite number of small classes (each containing only one *O*) at the bottom on up through classes which are fewer and fewer in number, but increasingly larger in size. At each level of the class hierarchy the classes are all mutually exclusive.

At each level the set of *O*s (or, rather, *θ*s because they are already occupants of the congitive space) which *S* has previously experienced is partitioned into classes. These classes, or parts of classes, can be combined with other classes, or parts of classes, on the same level to form higher level classes. For example, a *θ* may be a member of the class "cottage" which can be combined with the classes "apartment," "house," "mansion,"

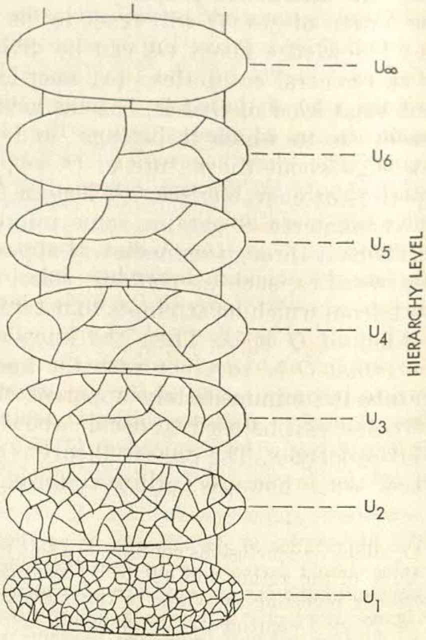


FIG. 1. The hypothetical structure of the cognitive space at various levels of the class hierarchy. (At the lowest level each class contains one *θ* and at the highest level one class contains all *θ*s.)

etc., and subsumed under the higher order class "residence." This in turn can be combined with "store," "depot," "bank," and the like, to form the yet higher order class "structure," and on and on.

Class Function. In the course of S's past experience with a θ he has experienced it as the member of a number of different classes. This is due to alternative names for the same class, errors in labeling, multiple-class memberships (such as a man belonging to religious, professional, and social groups), etc. As a result it is not always clear to which class a given θ belongs at the time the information is needed to aid an interaction, and thus a decision must be made about θ's membership.

The rule for determining which of the $k = a$ through i classes at a specific hierarchy level will be assigned to a particular θ is called the class function. The class function is in turn determined by two variables; the relative frequency with which θ and a particular class have been associated and whether or not the class is the one with which θ was last associated. These two variables are functions on the set of classes for each θ and their relative importance can be treated as weights on the functions.[2] For each θ the sum of these two weighted functions on the set of classes defines the class function. The class for which the sum is greatest should be the one assigned to the θ under consideration.

Inference within the Class Hierarchy. When one of θ's class memberships is known, its memberships at many of the higher levels in the hierarchy can be inferred on the basis of the proportion of θ's fellow class members known to belong to each of the higher order classes. Similarly, θ's lower order classes can be inferred from the proportion of θs from each of the lower order classes which are known to constitute the class to which it belongs. Because these proportions can be used to make such intrahierarchy inferences it is important to note that, because there are fewer classes at higher than at lower levels, a large proportion of O's known classes will usually go into one higher order class. On the other hand, O's known class is usually made up of a number of lower order classes. This means that inferences from O's known class *downward* in the heirarchy usually will be based on more rectangular distributions of proportions than will inferences made upward in the hierarchy. Consequently, downward inferences are frequently more uncertain than are upward inferences.

Selection of a Hierarchy Level. When an O is encountered in the environment S must select a level in the class hierarchy at which to attempt a class inference (i.e., he must select the set of classes to be considered for O's membership). While the lowest level possible would be best, because this permits intrahierarchy inferences of O's membership in a number of higher order classes, this is seldom possible. To classify an O as a member of a very low level class it is necessary to know a good deal about O's attributes, because lower level classes are more tightly defined than higher order classes, e.g., structure versus cottage, with the lower levels containing only a few very similar Os in each class. Because S seldom possesses much information at the moment he is forced to make his decision he usually must settle for a higher order class.

A second condition which tends to make S select a high level for the class inference is that the decision about the specific class is more likely to be correct at higher levels. (There is a smaller possibility of error in assigning an O to the broad class, structure, than in deciding it is a member of the smaller, more tightly defined class, cottage.) These two factors, lack of information about O and desire to be correct in classifying it, tend to make S select a high level in the class hierarchy for deciding O's class.

On the other hand, after S decides what O is, he usually intends to make further inferences about its cue values. These cue inferences are based on knowledge he has about *O*'s fellow class members. As a result, he does best to select a lower order class initially because such classes are smaller and are composed of more homogeneous *O*s.

---

[2] The importance of the last associated class or cue value should decrease as the time between the last encounter and the present encounter increases. This lends consistency to the assignment of classes and cues to the θs by insuring that previous inference errors, temporary changes in the O, etc., will have less influence as time goes on and that the highest relative frequency will exert more and more influence on the assignment.

In general then, the higher the order of the class the more likely is the identification of *O* to be correct; the lower the order of the class the more likely are subsequent cue inferences to be correct. The class level at which these two factors balance one another will be the one selected for the identification of an unfamiliar *O*. The mechanism for determining where these factors balance will be more closely examined later in the discussion.

Cue Dimensions

The classes are defined as attributes for which only two values exist: presence or absence. For some attributes, however, it is possible to discriminate among *O*s not only on the basis of the presence or absence of an attribute, but also on the basis of the *degree* to which the attribute is present. For example, some *O*s not only possess the attribute of weight but some are heavier or lighter than others. When such discriminations can be made among *O*s possessing an attribute it qualifies as a cue dimension and the degree to which each *O* possesses one of these attributes is *O*'s value on the cue dimension. (A rule of thumb for distinguishing cues and classes: Proper nouns, common nouns, collective nouns, and personal pronouns are classes while abstract nouns, adjectives, participles, and adverbs are usually cue values.)

Cue-Dimension Hierarchy

Just as classes exist in a hierarchy so too is there a hierarchy of cue dimensions. The cue-dimension hierarchy extends from the higher order, culturally defined attributes, like beauty or status, on down to the sensory level. Most of the meaningful cue values we associate with *O*s are only remotely related to the original sensory stimulation. Stimuli emanating from *O* stimulate *S*'s receptors and produce sensations, e.g., loudness, light, color, heat, etc. These lowest order cues are then used to infer slightly higher order cues such as edge, texture, surface, and size. In turn, these cues lead to the inference of even higher order cues such as square corners, planes of surface, orientation in space, extension, and so on. This procedure leads *S* up the cue-dimension hierarchy until he finally arrives at a point at which he has enough information both to infer *O*'s class and to place its cognitive representation (its θ) appropriately in the cognitive space. After this point, knowledge of both the class and the cue values determine subsequent inferences about the yet higher order cue values.

Cue Function. The method of assigning a cue value to a θ is parallel to that used for assigning classes. This rule, called the cue function, treats the cue values from one dimension as a set of elements which are to be assigned to each θ in the cognitive space. Because of temporary and permanent changes in *O* over time, variations in viewing conditions, etc., any θ may have had a number of associated values from the dimension. This history generates a relative frequency distribution for the cue values which, together with the last value associated with θ, determines the cue value to be attributed to θ. These variables can be treated as functions on the cue space and weighted for relative importance. The cue value for which the sum of the weighted values of these two factors is greatest is the value which *S* should assign to the θ under consideration.

Class and the Cue-Dimension Hierarchies. The similarity between the class heirarchy and the cue-dimension heirarchy is not so great as it at first appears. The class hierarchy, as illustrated in Figure 1, is composed of sets of classes into which θ are repartitioned at every level of the hierarchy. Knowledge of its class at one level permits intrahierarchy inferences about θ's class memberships at other levels of the hierarchy. On the other hand, the cue hierarchy does not consist of the reassignment of elements at each level. The levels of this hierarchy only indicate the abstractness or "remoteness" (Brunswik, 1956) of the various cue dimensions. The θ's value on each succeeding higher order cue dimension must be inferred from the lower order cue values.

COGNITIVE SPACE

The assumed structure of the cognitive space is similar to that of analogous concepts proposed by Kelly (1955); Osgood, Suci, and Tannenbaum (1957); Overall and Williams (1961); Sarbin, Taft, and Bailey (1960), and others. It is a multidimensional space which is defined by the cue dimensions to which *S* attends.

The cognitive space provides an information storage and retrieval system for *O*s which *S* encounters. Each cue value associated with an *O* can be located on the appropriate cue dimension bounding the space. All of these values define a particular point, called θ, which is unique to *O* and which is the key both to retention of information about *O* and to retrieval of this knowledge when it is needed. The θ serves to retain knowledge through its location because its location is a summary of everything *S* knows about *O*'s cue values. It serves in the retrieval of knowledge about cue values because specification of its location designates the appropriate θ and permits the determination of its cue values through the cue function. It serves in the retrieval of knowledge about *O*'s class membership because θ is linked to the classes by the class function.

INFERENCE OF CLASS MEMBERSHIP

Recognition

When an *O* appears before *S*, only a few lower level cues are apparent. Using these cues, *S*'s first task is to decide whether or not this *O* has ever been experienced before, i.e., if it is familiar or unfamiliar. If it is recognized as familiar, a great deal may already be known about it and this knowledge can be utilized in the ensuing interaction. Because of changes in situational conditions, however, as well as changes in an *O* over time, the immediately apparent cue values associated with an *O* may be slightly different from one encounter to the next. The *S*'s problem is to decide whether the cue values associated with a θ in the space are sufficiently similar to those associated with the present *O* to permit the assumption that θ derives from a previous encounter with this same *O*.

When an *O* is encountered the first step in the inference process is to determine at which level of the class hierarchy the inference will be made. That is, *S* must select the set of classes which he will consider as possible candidates for *O*'s membership. The selection is determined by *S*'s desire to be correct in his class inference, his desire to be correct in his subsequent cue inference (cf. section on Selection of a Hierarchy Level), and the context in which the inference is being made.

The second step in the inference process is to use *O*'s immediately apparent cue values to tentatively define a cell in the cognitive space, called ϕ. Next, the distance is measured between ϕ and every θ in the cognitive space.[3] The smaller the distance between ϕ and the nearest θ, the more likely it is that they both derive from the same *O*. There is, however, some critical distance beyond which *S* will reject the hypothesis that both ϕ and the nearest θ derive from the same *O* and therefore that *O* is not recognized. (See Ames, 1949, and Arnoult, 1956, concerning *Ss*' assumption of identity between similar *Os*.) The size of the critical distance is determined by *S*'s motivation to be correct in his inferences. If it is very important to be correct, the size of the critical distance will be very small and the nearest θ will have to be nearly identical to ϕ to be accepted. If accuracy is less important greater dissimilarity can be tolerated.[4]

---

[3] No specific method has been given for measuring distances in the cognitive space because so many alternatives are available and, for research, the final choice would depend on the conditions of the experimental situation. One method might be to use D^2 (Cronback & Gleser, 1953; Osgood & Suci, 1952) computed for the cue dimensions common to ϕ and each θ. Similarly, variations of the scaling procedures developed for decision theory could be used, and the scaling techniques similar to that developed by Ekman, Goude, and Waern (1961) might also be of use.

The decision paradigm, as well as the concept of the critical value fluctuating as a function of motivation, closely resembles statistical decision theory and is based on the Tanner and Swets (1954) approach to stimulus detection.

[4] When a number of θs are all the same distance from ϕ and all lie within the critical range around it, *Ss* apparently match the relative frequency of their choices of each θ to the relative frequency with which they have previously seen in each θ, i.e., probability matching (Binder & Feldman, 1960). In the laboratory *S* is not permitted to ob-

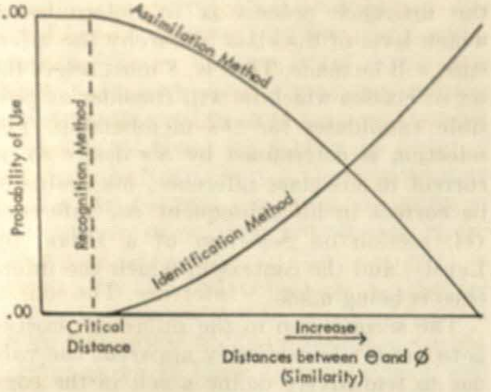


FIG. 2. Probability of use of each inference method as a function of the distance between θ and ϕ.

Another determinant of whether or not one θ will be accepted as having derived from O in question is S's "familiarity" with the θ (Arnoult, 1956; Nobel, 1954). If one θ is more familiar than another, even though their distances from ϕ are equal, S should be more inclined to accept the former as having derived from the present O.

Familiarity is primarily determined by the frequency with which θ has been encountered. However, it is somewhat more than just frequency, it is also influenced by the duration of the encounters and how much S knows about θ's various class memberships and cue values. In some cases, of course, this reduces primarily to frequency of past encounters, and in these cases the influence of frequency of past encounters upon recognition is well documented by a sizable body of literature (e.g., Arnoult, 1956; Bruner, 1957; King-Ellison & Jenkins, 1954; Postman & Rosenzweig, 1956; Solomon & Postman, 1952). However, a more general definition of familiarity is needed for our present purposes. Here *familiarity* is defined as the sum of the number of times a θ has been classified and the number of cue values known for it. Thus, while familiarity is closely related to relative frequency of encounter, it also takes total knowledge into account.

When it is decided that ϕ lies close enough to a θ to assume that they derive from the same O, the class function serves to assign a class to θ and, in turn, to O. Then S can assume O is a member of the class and proceed with the interaction. If θ has never before been assigned a class on that class hierarchy level, S has two courses of action open to him. First, he can select a higher or lower hierarchy level and attempt to assign one of the classes on the new level. Then he can make intrahierarchy inferences up or down to the level at which he previously wanted to make the inference. Second, if the intrahierarchy inference procedure is too involved or does not yield sufficiently certain inferences, S can fall back upon the identification method which will be explained below.

Assimilation

When the distance between ϕ and the nearest θ in the cognitive space is greater than the critical value, S can assume that they do not derive from the same O. However the matter does not end there. If ϕ and θ are reasonably similar, it is prudent to consider whether this similarity is not paralleled by a similarity in class membership and in as yet uninferred cue values. Thus, even though O is probably not the same one which has been encountered before, θ's class still may be appropriate because θ and ϕ are so similar in their other aspects. When S can clearly discriminate between θ and ϕ and yet he adopts θ's class for ϕ, and therefore O, the process is called assimilation.

The hypothetical curves in Figure 2 describe the effect of the distance between ϕ and the most similar θ upon S's choice of an inference method. First, when familiarity is held high and constant, up to the critical distance, ϕ and θ are assumed to derive from the same O and the present O is recognized as belonging to the class associated with θ. Just beyond the critical distance ϕ and θ are still extremely similar and even though S does not recognize O he is still likely to assimilate θ's class for O. As the distance increases, the adoption of the most similar θ's class for O becomes less likely until at large distances it becomes extremely unlikely, and at the same time the probability in-

---

tain additional cues. Ordinarily, however, another cue value would be sought in order to break the deadlock.

creases that S will use the identification method which will be discussed presently.

In Figure 3 are presented the hypothetical relationships between familiarity with θ and the probability of S using recognition, assimilation, or identification to make his inference. These two curves are from a family of curves which change as a function of the difference between ϕ and θ defined in Figure 2. If one placed the right margin of Figure 3 against the left margin of Figure 2, the result would be a three-dimensional graph of the family of curves.

When the distance between θ and ϕ is held small and constant, an increase in S's familiarity with θ results in an increase in the probability that S will use assimilation or recognition to infer O's class memberships, and a decrease in the probability that he will use identification. On the other hand, a larger distance between θ and ϕ results in a lower assimilation-recognition curve and a higher identification curve with corresponding changes in probabilities for any given frequency values.

Identification

When an O's cue values define a ϕ which is too different from the nearest θ for either recognition or assimilation, S must obtain the class through identification. Unlike assimilation, identification is based upon a large number of θs which are similar to O, rather than upon just one. Inferences based on a large sample of similar θs are more likely to be correct than those based on just one previously experienced θ. So, when S's inferences for unfamiliar Os require a high degree of accuracy, assimilation is likely to be abandoned and identification used to infer Os' class.

The mechanics of the identification process were outlined in the previous monograph. However, to refresh the reader's memory it will be recalled that the identification process, like the recognition and assimilation processes depends upon the immediately apparent cue values associated with O to be identified. Each of these values has been experienced by S in the course of his past experience with other Os, Os which were members of classes and which are represented by θs in the cognitive space. Given a specific level of the class hierarchy at which to work, *each* cue value associated with O defines a relative frequency distribution composed of θs which possess the cue value, d, and which are members of each of the different classes, $k = a$ through i. For each cue value the proportion of θs in each class, written $P(k/d)$, constitutes $1/n^{th}$ of the evidence for the class in question being the best choice for S's inference. The sum of the $\frac{P(k/d)}{n}$ over all of O's known cue values for each of the $k = a$ through i classes gives the proportion of the total evidence yielded by the n cue values that indicates that each class is the best bet for O's membership.

$$E(k) = \sum_{d=1}^{n} \frac{P(k/d)}{n}. \qquad [1]$$

Where,

k = a specific class under consideration as a possibility for O's identity; $k = a$ through i

d = a dimension on which O's cue value is known; $d = 1$ through n

The S should choose the class with the highest $E(k)$ for his inference about O's class. When this inference has been made ϕ takes on the class as an attribute and becomes a permanent θ in the cognitive space.

Some Properties of the Equation. The simple additive form of Equation 1 was deliberately selected because there is some evidence that it may be sufficient and because, while it is possible that other, more complex

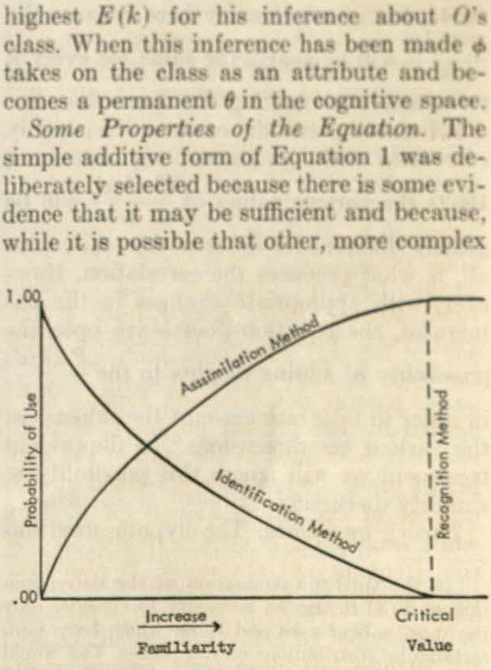


FIG. 3. Probability of the use of each inference method as a function of familiarity with the θ.

forms will prove to be more adequate, it is probably best to start with the simplest assumptions and then complicate the theory where need be. (Anderson, 1961, successfully used a similar additive equation to predict *Ss*' impressions of *O*s based on descriptive adjectives.)

Equation 1 has two particularly interesting properties. First, if any of the $\frac{P\ (k/d)}{n} = 0$, the equation does not reduce to $E(k) = 0$ as it would if the equation were multiplicative. This seems to agree with common sense. Just because the cue value has never before been associated with a member of a particular class does not mean that it is impossible that the present *O* belongs to the class; this may just be the first time *S* has encountered this situation.

A second interesting property of the equation is that it is not necessary to weight the various $\frac{P\ (k/d)}{n}$ according to the ability of their corresponding cue dimensions to discriminate among the classes. Such a weighting procedure would be used in a correlational analysis (Hoffman, 1960; Todd, 1954), but is unnecessary here because the $\frac{P\ (k/d)}{n}$ will be nearly the same for every k for a nondiscriminating cue dimension. For discriminating cue dimensions (i.e., highly correlated with the classes, e.g., Summers, 1962) the various values of $\frac{P\ (k/d)}{n}$ will be grossly different for different k's; this, after all, is what produces the correlation. However, with appropriate changes in the numerator, the equation does leave open the possibility of adding weights to the $\frac{P\ (k/d)}{n}$ in order to take into account the salience of the various cue dimensions.[5] In the present treatment we will ignore this possibility to simplify discussion.

Theory and Data. The hypothesized nature of the class inference methods and the contents of Figure 2 are based upon the results of the study reported in the preceding monograph. In that study it was found that in the early stages of training, when *O*s were unfamiliar to *Ss*, inferences corresponded to the identification method's maximum $E(k)$. As training progressed, and *O*s became more familiar, *Ss*' inferences for confirming cases continued to correspond to the maximum $E(k)$. (For these cases, of course, recognition, assimilation, and identification all lead to the same response). For infirming cases, however, *Ss*' inferences changed from the (incorrect) maximum $E(k)$ to the correct class as familiarity with *O*s increased.

For the test cases, on the other hand, recognition could not be used; *Ss* had never seen the specific cue combinations which comprised *O*s. However, *Ss* could use assimilation for *O*s which were similar to *O*s experienced during training. And, indeed, it was found that *Ss* used the identification method, i.e., assigned a test *O* to the high $E(k)$ class, only when the test *O* was not similar to any *O* in the training set. When the test *O*'s cue values are similar to those possessed by a familiar *O* (either confirming or infirming), *Ss*' inference for the test *O*'s class was the one associated with the familiar *O*, i.e., the class was assimilated from the familiar *O*.

These results have determined the form of the present theory. The test *O* results indicate that while *Ss* use cue probabilism for unfamiliar *O*s, they tend to depend heavily upon the similarity between previously experienced *O*s and *O* for which the inference is to be made in order to reach a decision about the latter's class. The infirming case results imply that before recognition or assimilation can take place it is necessary for the old *O* to have been experienced a number of times and that before this familiarity is attained *Ss* utilize cue probabilism in order to make their inferences.

Apparently *Ss* first attempt to infer an *O*'s class by the recognition method, accepting only familiar θs as possible candidates. If no familiar θ is sufficiently similar to ϕ to permit recognition, *S* resorts to assimilation—still requiring that a familiar θ be the basis of the assimilation. If there is still no sufficiently familiar θ, *S* uses the identification method which apparently utilizes every θ in the cognitive space rather than just the most frequently experienced.

---

[5] In the further examination of the determination of $E(k)$ it may be necessary to consider only the most salient cues and those which have nonrectangular distributions on the classes. This would limit the size of n and thereby prevent the values of all of the $E(k)$ from becoming indiscriminably small.

INFERENCE OF CUE VALUES

When an *O* is encountered, it can be considered the possessor of two kinds of cue values, known and unknown. The known cues include all of the immediately apparent aspects of *O* (which are used to make inferences at the lower levels of the cue dimension hierarchy and which form the basis for the inference about *O*'s class) and all of the cue values which the cue function assigns to *O*'s θ in the cognitive space. All others are designated unknown cue values. The most important unknown cue values are the higher order ones upon which *S*s' interactions with *O*s are based, e.g., intelligence, value, use, and friendliness. These higher order cue values, more so than an *O*'s class, are the end point in *S*'s search for knowledge about *O*. It is only by knowing these attributes of *O* that *S* can properly plan his interaction and attain his long range goals.

Recognition and Cue Values

Just as recognition of an *O* reveals its class membership, so too does it reveal many of *O*'s cue values which, while known from past experience, are not apparent at the moment. These unknown cue values are associated with θ in the cognitive space and are available for use when the proper θ is located. In this way recognition permits the utilization of past experience in the planning and execution of interaction with *O*. The cue values associated with θ are, however, only those which past experience has provided. If other, as yet unknown, cue values are required for the interaction, *S* must fall back on either assimilation or identification in order to infer them.

Assimilation and Cue Values

When an *O* is not recognized, or when specific attributes of a recognized *O* have never been experienced, *S* can utilize assimilation to provide the unknown cue values. The procedure is the same as for an assimilation of classes; the needed cue values can be assimilated from the most similar θ in the cognitive space.

Identification and Cue Values

When neither recognition nor assimilation provides the required cue values for an *O*, they can still be inferred in much the same way that classes are identified. When the cue inference is begun *O*'s class is already known through recognition, assimilation, or identification. Moreover, some of its lower order cue values are known and sometimes, if it has been recognized or if cue values from other θs have been assimilated, some of its higher order cue values are also known. Knowledge of the class together with each of the known cue values provides the foundation for inferring the necessary cue values.

In terms of the cognitive space, *O* possesses both known cue values and a known class; the goal is to infer a cue value on a specific cue dimension for which the value is unknown. The reasoning underlying this inference is parallel to that upon which the class identification method is based. Here, however, *S* need not deal with all θs in his cognitive space. He can begin by considering only θs that belong to *O*'s class as relevant to the inference. Then each of *O*'s cue values defines a relative frequency distribution upon the unknown cue dimension. This distribution yields the proportion of θs (*a*) that belong to *O*'s class, k; (*b*) that possess the same cue value as *O* possesses, d; and (*c*) that possess each of the possible values of the unknown cue dimensions, $c = 1$ through i. This proportion, $P(c/k,d)$, indicates the likelihood that a value c on the unknown cue dimension is the best bet for the inference given *O*'s class. The cue value associated with the highest proportion is the best bet for the inference about *O*'s unknown cue value.

When, as is usually the case, *S* knows more than one of *O*'s cue values there may be conflicting inferences yielded by the different distributions. As in the inference of classes it is possible to compromise among these conflicting inferences by summing, for each unknown cue value, the proportions contributed by each known cue value (keeping in mind that each proportion is only $1/n^{th}$ of the evidence for that specific value of the unknown dimension).

$$E(c) = \sum_{d=1}^{n} \frac{P(c/k,d)}{n}. \qquad [2a]$$

Where,

c = a cue value under consideration as O's unknown cue; $c = 1$ through i

d = a dimension upon which O's cue value is known; $d = 1$ through n

k = O's class

n = the total number of dimensions upon which O's cue values are known

$E(c)$ is the total evidence indicating that each cue value, c, from the unknown cue dimension is the best bet for O's cue value. The S's inference should be the cue value associated with the highest value of $E(c)$.

Lower Level Cue Inferences. For inferences of the lower level cue values, those inferred from sensation and used to infer classes, Equation 2a must be modified. At this stage S knows only the very basic sensation cues—which are multivalued dimensions at the level of the nerve if not at the level of the neuron—and perhaps a few other lower level cue values which have already been inferred. The class is not known and therefore it must be dropped from Equation 2a, which becomes:

$$E(c) = \sum_{d=1}^{n} \frac{P(c/d)}{n}. \qquad [2b]$$

Through the inference process described by Equation 2b O's basic cue values are built up and stored in the cognitive space in terms of the location of ϕ. After enough values are known, or when no more are forthcoming, O's class is determined through recognition, assimilation, or identification. Then, higher order cues can be inferred through their association with the recognized θ, through their assimilation from a similar θ, or, now that the class is known, through identification inferences and Equation 2a.

Relevant Research. When an O is recognized, its cue values are immediately revealed for a large number of cue dimensions. This is a familiar and obvious happening which is exactly what we mean in the everyday use of the word "recognize." For assimilation and identification, however, the cue inference process is less apparent and as yet, little appropriate research has been done.

Some light is cast upon the assimilation and identification process by data obtained by Beach (1961).[6] This research followed the same plan as that described in the preceding monograph. The Ss were trained on a deck of cards and then were required to make inferences about the values of cues which were missing from test cards. In most cases, the inferences about the unknown cue values were dictated by assimilation; the cue value possessed by a training O which possessed cues similar to the test O's known cues was given as the test O's missing cue. In some cases, however, identification played a role and the inference was predicted by Equation 2a. Which method best predicted Ss' inference for a given test O appeared to be governed by the degree of dissimilarity between the test O and the most similar O from the training set. While sketchy, these results imply that assimilation and identification play a role in Ss' inferences of unknown cue values.

There are two other studies which are relevant to the role of the identification process in cue inference. In the first of these (Goodnow, 1954) Ss were taught to classify Os on the basis of three dichotomous cue dimensions and then were shown test Os with one of the cues missing and were asked to infer the missing value. It was found that Os' known cues first were used to determine their classes and then the classes were used to determine the inferences about the missing cues. Unfortunately, the results do not yield information about whether or not Os' known cues influenced the inference in the manner described by Equation 2a or whether the inferences were determined by the class alone. Nevertheless, these results indicate that inference of O's class is a first step in the inference of unknown cue values and that these unknown cue inferences are influenced by *S*'s knowledge of *O*'s class membership.

---

[6] Note that both the class and cue inference equations in Beach (1961) have been rejected in favor of those given above.

In the second study (Bolles & Bailey, 1956) five *S*s were told 54 *O*s' class memberships together with some of their cue values and asked to infer *O*s' sizes. The average correlation between the inferences and *O*s' actual sizes was .994. Moreover, *S*s' errors were larger for the less familiar classes of *O*s and for the classes which had large size variance among their members. Bolles and Bailey interpret these results as showing the importance of prior knowledge of *O*'s class upon unknown cue inferences. The additional cues given about *O*s are confounded, however, with the knowledge of the class. Consequently, the relative roles of these two factors are not known, and it is impossible to tell whether or not Equation 2a could account for the results.

The results of these three studies are much too vague to be seriously considered as evidence for the theory. At present, research is being undertaken which will attempt to examine the cue inference processes more thoroughly.

INFERENTIAL ACCURACY

Revision Process

It is important for *S*'s inferences about an *O* to be as accurate as possible. This is true because the inferences must stand up when they are used in interactions with *O* and because they become associated with θ in the cognitive space and form the basis for inferences (through assimilation and identification) about other *O*s. To assure the utmost accuracy for the class and cue values associated with a θ each inference is constantly being re-evaluated in light of information gained in the course of interaction with *O*.

The revision process consists of utilizing environmentally derived cues to reinfer each attribute of θ by the identification method. The new inference is then compared with the old one, no matter which method originally produced the latter.

Affirmation of the Old Inference. When the old and the new inferences agree, indicating the stability of *O*'s attribute and of the accuracy of the previous inference, the inference is retained as an attribute of *O* and the fact that it was affirmed is recorded for future reference. The recording of a class or cue value's affirmation is accomplished by *affirmation frequency distributions* on the set of classes at each level in the hierarchy and upon the sets of cue values from each dimension in the cognitive space. These distributions record the frequency with which each class and each cue value has been experienced as an attribute of θ. Affirmation frequency is an important concept because it serves to tie together many of the previously introduced mechanisms. Therefore, before proceeding with the discussion of its role in revision let us look at its relationship to these other parts of the theory.

Affirmation frequency, through the appropriate functions, plays a role in the assignment of classes and cue values to θs in the space. As was discussed earlier (see sections on Class Hierarchy and Cue Function) assignment is dependent partially upon the relative frequency with which the classes or cues have been associated with θ. These relative frequencies are derived from the affirmation frequencies for the appropriate set of classes or cues. Also, in the discussion of the factors which influence the selection of a θ for recognition or assimilation (see p. 26) it was stated that one important determinant was *S*'s familiarity with θ. Familiarity is determined by the sum of the affirmation frequencies across the set of all classes together with the total number of cue values known for θ (from the functions which assign the last associated cue value to θ). In this way affirmation frequency plays a role in determining which inference method will be used to infer an *O*'s class.

Revision of the Old Inference. When a θ's attribute is re-evaluated and affirmed, it is retained as one of the classes or cue values and its frequency of affirmation is increased. If, however, the new inference disagrees with the old one, a decision must be made about which is correct. If it is decided that the old inference is incorrect, another tally must be made for the affirmation distribution on the appropriate set of classes or cue values, and the function which records the last associated class or cue value must be revised. When the latter change is made for cue values, the result is a change in θ's location in the cognitive space.[7]

The revision decision is made by comparing two factors, the frequency with which the old inference has been affirmed and the magnitude of the $E(k)$ or $E(c)$ for the new inference. Obviously, if the old inference has been frequently affirmed, it is not prudent to replace it with the new inference unless the latter is shown to be very probably in light of new information about θ. On the other hand, an old value which has been affirmed only a few times is not necessarily a better choice than the new inference.

Inferential Confidence

To compare frequency of affirmation and the magnitude of $E(k)$ or $E(c)$ it is necessary to introduce another concept into which they both can be transformed. This concept is degree of inferential confidence. It is positively related to both frequency of affirmation and to magnitude of $E(k)$ and $E(c)$. The inference with the higher confidence value is the one assumed to be correct for θ.

Measurement of Inferential Confidence. Like affirmation the concept of inferential confidence ties together many of the mechanisms which have been introduced throughout the previous discussion. It is defined as the degree to which *S* believes that a decision or an inference is accurate, and it can be measured in a number of ways. The most candid method, and one with apparently high validity, is merely to ask *S* how confident he feels about a decision or inference (see, for example, Anderson & Walen, 1960; Irwin, 1953; Pollack & Decker, 1958). A more complex method is to ask *S* to place bets on his decisions or inferences (Edwards, 1961). Another method, which to the author's knowledge has never been used, is to utilize the degree of specificity or generality in *S*'s interactions with an *O* to determine his confidence in the inferences upon which the interaction is based. If *S* is confident about the accuracy of his inferences about an *O*, those aspects of the interaction which are based on the inferred attributes should reflect the confidence. In this situation, high inferential confidence should be revealed in the form of fairly specific predictions about *O* and *O*'s reaction to *S*'s behavior toward it. If, on the other hand, *S* lacks confidence in his inferences about the relevant attributes of *O*, he can make more cautious predictions about *O*. Less specific, general predictions will permit interaction—and thus new knowledge with which to re-evaluate the questionable inferences—without jeopardizing *S*'s goals through a *faux pas*. This sort of behavior is frequent in social situations in which *S*'s interaction with *O* is dictated solely by social form until he has an opportunity to "size-up" *O*. Proper examination of the generality and specificity of behavioral interactions with *O*s could possibly lead to a method of estimating *S*'s confidence in the inferences underlying the interaction.

Inferential Confidence and Inference Method. Inferential confidence is essentially a scale of subjective probabilities about the correctness of an inference. Each inference is associated with a value from this scale, and the critical value for acceptance or rejection of the inference is determined by *S*'s motivation to be correct. Thus, for example, in Figure 2 the critical distance is the degree of difference between θ and ϕ beyond which *S*'s inferential confidence is too low to accept that they derive from the same *O*. Similarly, the rejection of assimilation and the switch to identification is governed by *S*'s confidence about assimilation inferences when large distances exist between θ and ϕ or when the frequency of contact with θ is small.

---

[7] In light of the present analysis it must be assumed that *S*s in the experiment reported in the previous monograph made two inferences for the cards. The first inference utilized cues on the faces of the cards and resulted in correct inferences for the confirming cards and in incorrect inferences for infirming cards. The second inference used these cues and the answer given by the experimenter—which, by definition, was a perfectly valid cue with a $P(k/d) = 1.00$ for the correct class. For the infirming cards the second inference disagreed with the first and thus, through revision, replaced the former as the card's class in *S*'s cognitive space. However, until familiarity with the cards had progressed to the point that he could utilize regonition, *S* continued to give the incorrect inference based on the cues on the faces of the infirming cards. As he began to utilize recognition, the inferences were the classes stored in the cognitive space and, for infirming cards, ceased to be the high $E(k)$ class and began to be the correct class.

Inferential Confidence and the Class Hierarchy. Inferential confidence also plays an important role in the selection of the level in the class hierarchy where an *O*'s class will be inferred. At the lower levels of the hierarchy the classes are more tightly defined and contain very few θs. As a result, assignment of *O*s to classes on these levels runs a good chance of error. However, if the $E(k)$ for the most likely class on one level is not large enough to warrant sufficient inferential confidence about being correct, *S* can move up the hierarchy until a satisfactory level is found. While *S*'s motivation to make a correct class inference will tend to make him select a high level in the hierarchy at which to work, his motivation to be correct on subsequent cue inferences will tend to make him select a low level. The hierarchy level at which the class inference is finally made will be where *S*'s degree of confidence for the class inference balances the degree of confidence for subsequent cue inferences (cf. section on Selection of a Hierarchy Level).

As these examples indicate, inferential confidence provides a common link for many of the concepts introduced throughout the theory. In addition to all of this, however, inferential confidence also governs another aspect of *S*'s inference behavior—the search for new information about *O*.

Cue Search

When inferential confidence is below the degree required by the existing level of motivation and the inference is assumed to be of questionable validity, some sort of action is required to remedy the situation. The action usually consists of a search for further information upon which the inference can be re-evaluated and either affirmed or revised. This search for information, called cue search (Bruner, Goodnow, & Austin, 1956), may be either active probing of the environment through questions, manipulation, reading, or the like, or it may be passive reception of information in the course of cautious interaction with *O*. In either case, it consists of behavior in which *S* obtains totally new information about an *O* or in which he again receives information he already possessed about *O*. The former is used to expand his knowledge about *O* and to re-evaluate, and perhaps revise, previous inferences. The latter adds to the various affirmation frequencies and raises *S*'s inferential confidence about those particular attributes of the *O*.

If *S* is not wholly confident that he recognizes an *O*, he can search for information about *O*. In this situation cue search will be initiated when the distance between θ and ϕ is very near the critical distance. If *O* is indeed familiar to *S*, the additional information will usually decrease the size of the distance and thereby permit recognition. If it is unfamiliar, the distance will usually increase and *O* will not be recognized, and either assimilation or identification will be used to make inference about *O*.

Inferences based upon identification will lead to cue search if no $E(k)$ or $E(c)$ is sufficiently high to warrant confidence in the accuracy of the inference. Indeed, whenever a number of inference alternatives are approximately equally likely, whether they are based on recognition, assimilation, or identification, cue search will usually be undertaken to break the deadlock.

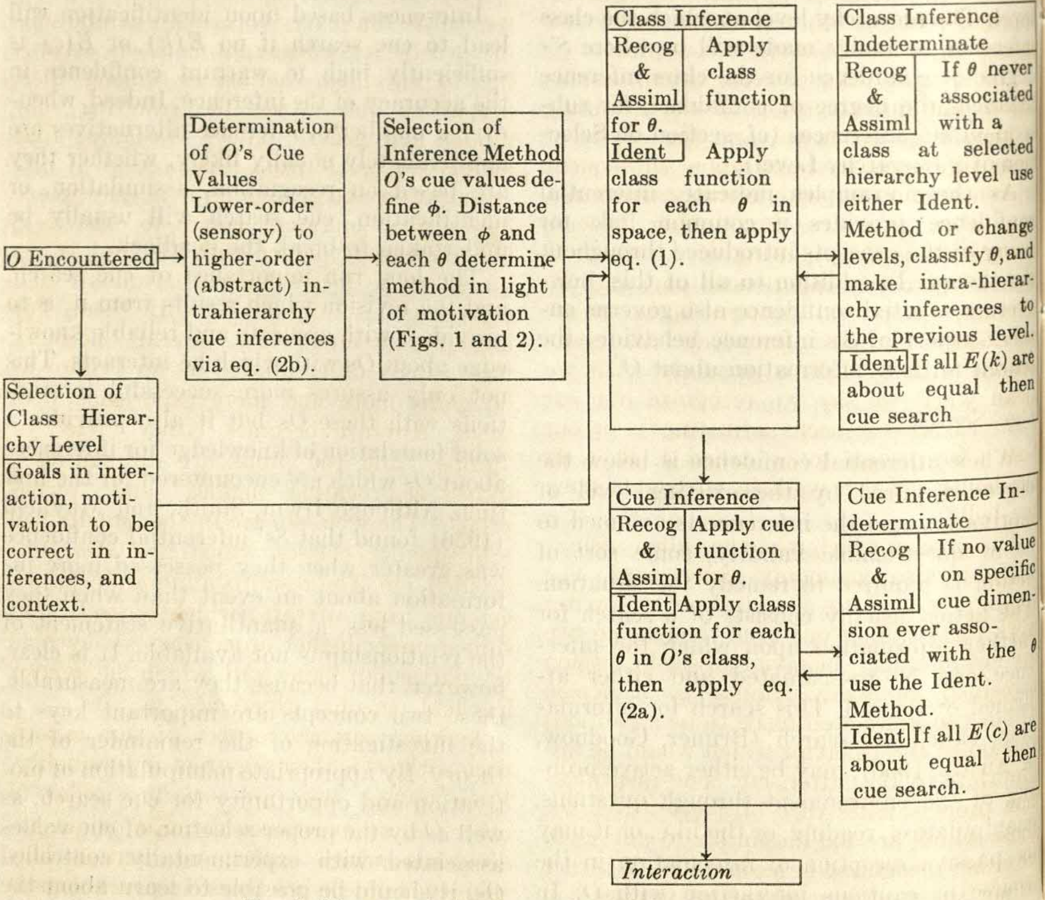
The long run importance of cue search, and the revision which results from it, is to provide *S* with accurate and reliable knowledge about *O*s with which he interacts. This not only assures more successful interactions with these *O*s but it also provides a solid foundation of knowledge for inferences about *O*s which are encountered for the first time. Although Irwin, Smith, and Mayfield (1956) found that *S*s' inferential confidence was greater when they possessed more information about an event than when they possessed less, a quantitative statement of the relationship is not available. It is clear, however, that because they are measurable, these two concepts are important keys to the investigation of the remainder of the theory. By appropriate manipulation of motivation and opportunity for cue search, as well as by the proper selection of cue values associated with experimentally controlled *O*s, it should be possible to learn about the empirical relationships which are assumed to exist between these and the other theoretical concepts.

OVERVIEW AND DISCUSSION

The major points of the theory and the hypothesized temporal relationships are outlined in Figure 4.

First, O is encountered in the environment and at the same time the class hierarchy level is determined on the basis of S's goals, his motivation to make a correct class inference and subsequent cue inferences, and the context of the encounter. Next, through Equation 2a, O's immediately apparent attributes establish its cue values on the cue dimensions. These cue values define the position of ϕ within the cognitive space. The ϕ is compared to θs, and a decision is made about whether or not it is similar enough to any θ to permit the assumption that they both derive from the same O. If the similarity is sufficient, the recognition method is adopted. If similarity is high but is not sufficient to warrant recognition, the assimilation method is adopted. If the similarity between ϕ and each θ is quite low the identification method is used.

Given the inference method and the hierarchy level, the next step is to infer O's class. For recognition and assimilation this is accomplished through application of the class function; for identification Equation 1 is applied. If there is no class to assign to the recognized or assimilated θ, S can change hierarchy levels, assign a class, and then make intrahierarchy inferences to the previously desired level. If this roundabout method does not lead to sufficient inferential confidence S can utilize the identification method for θ. If the identification method



O Encountered
Selection of Class Hierarchy Level
Goals in interaction, motivation to be correct in inferences, and context.
Determination of O's Cue Values
Lower-order (sensory) to higher-order (abstract) intrahierarchy cue inferences via eq. (2b).
Selection of Inference Method
O's cue values define ϕ. Distance between ϕ and each θ determine method in light of motivation (Figs. 1 and 2).
Class Inference
Recog & Assiml
Apply class function for θ.
Ident
Apply class function for each θ in space, then apply eq. (1).
Class Inference Indeterminate
Recog & Assiml
If θ never associated with a class at selected hierarchy level use either Ident. Method or change levels, classify θ, and make intra-hierarchy inferences to the previous level.
Ident
If all $E(k)$ are about equal then cue search
Cue Inference
Recog & Assiml
Apply cue function for θ.
Ident
Apply class function for each θ in O's class, then apply eq. (2a).
Cue Inference Indeterminate
Recog & Assiml
If no value on specific cue dimension ever associated with the θ use the Ident. Method.
Ident
If all $E(c)$ are about equal then cue search.
Interaction

FIG. 4. Diagram of the class and cue inference process.

does not produce an $E(k)$ for one class which is sufficiently greater than the $E(k)$s for the other classes, S must search for more information about O.

After O's class is determined S can move on to inferences about its unknown cue values. If O is recognized, or if its class is assimilated, θ may have the desired cue value associated with it. If there is no value associated with the selected θ, or if identification has been used to determine O's class, the unknown cue values can be inferred through Equation 2a. If the identification method fails to provide a definitive answer for the cue inference, then S must search for further information about O.

The theory is an attempt to build upon the research results reported in the previous monograph and to fit these results into a larger conceptual framework. However, care has been taken to make the general approach conform to a common sense view of how inference behavior takes place.

The specific roles of recognition, assimilation, and identification in the attaining of knowledge about Os, and the effects of these inference methods upon subsequent inferences, fit well with common-sense notions about cognitive functioning. Indeed, recognition is the equating of a present O with a previously experienced one, and the common behavioral pattern is to utilize previously obtained knowledge in interactions with O. This is precisely what we mean when we say we recognize our car ("the carburetor needs adjusting"), or our boss ("he prefers to be called J. R."), or our house ("I can walk in without ringing the bell").

The process which we have called assimilation is also commonly experienced. The techniques used to operate one typewriter work for a similar one even though the two are not identical; behavior toward a person perceived as highly similar to a person whom you know well tends to be a cautious version of the behavior which would be appropriate to the latter, etc. The identification process takes care of the leftover Os, those which are too dissimilar to any θs to be either recognized or assimilated. Whether or not the specific mechanisms proposed in the theory are entirely correct, these three kinds of inferences make sense and apparently cover the field.

The cue and class hierarchies are also common sense. When it is recognized that Os must be "reconstructed" from sense data, so to speak, the hierarchical structure of the cues becomes a necessity. It is clear that there are very real differences among cues; some are closely akin to sensory events while others are more abstract. Granted that cues all derive from sensation rather than from innate ideas, the abstract cues must be generated by the sensory events, and they in turn generate other knowledge about Os. It is equally clear that what an O is influences our interpretation of its properties; therefore, the O's class must be known before many of its attributes can be properly evaluated.

The class hierarchy is common sense if we realize that every class of Os can be subdivided by segregating its members according to specified similarities among their properties and giving the resulting subgroups new generic names. Assuming that all classes result from this process of differentiation and renaming, it is easy to see how a class hierarchy comes to exist as S has increased contact with Os and is required to be increasingly discriminate and correct in his interactions with them. For example, the broad class "aircraft" becomes increasingly differentiated by an aviator as he gains more and more experience with them. Aircraft becomes either "jets" or "props," and the props are differentiated into "reciprocating engines" and "turboprops." These classes are further subdivided into smaller classes which are also subdivided, etc. The result is a hierarchy which extends from aircraft (which is itself subsumed under higher order classes) down to the particular planes with which the aviator has had personal experience.

Another common aspect of our experience with Os is their apparent individuality. Even though they are usually similar to other Os, we experience them as unique individuals which differ, no matter how slightly, from their fellows. In psychological theorizing, on the other hand, there is a tendency to forget this fact. In the present theory, an effort has been made to acknowledge and to utilize an *O*'s individuality as an important part of the inference process.

Individuality is reflected in recognition and in its temporal priority in the class inference process. Recognition is the quintessence of dependence upon an *O*'s individuality and constancy over time. Similarly, cue inference through the identification process emphasizes *O*'s individuality by stressing the role played both by *O*'s class and by its unique cue values. The *O* is not merely relegated to a class and, once there, merged into the mass of *O*s which are its members. Rather, inferences about its cues are based on only those fellow class members which share its unique cue values.

The integrity of the individual *O* also is reflected in the theory's emphasis on affirmation and revision of knowledge about each of them. In the course of interaction with an *O* its eccentricities and peculiarities become known and old incorrect knowledge is revised. In this way *S* keeps abreast of changes in *O*s which populate his environment and, consequently, is capable of conducting successful interactions with both the old and the new *O*s which he encounters.

REFERENCES

Ames, A. *Nature and origin of perception*. Hanover, N. H.: Hanover Institute, 1949. Cited in C. E. Osgood, *Method and theory in experimental psychology*. N. Y.: Oxford Univer. Press, 1956.

Anderson, N. H. Application of an additive model to impression formation. *Science*, 1962, **138,** 817–818.

Anderson, N. H., & Walen, R. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60,** 111–120.

Arnoult, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51,** 269–276.

Beach, L. R. A mathematical theory of inference behavior. Unpublished doctoral dissertation, University of Colorado, 1961.

Binder, A., & Feldman, S. E. The effects of experimentally controlled experience upon recognition responses. *Psychol. Monogr.*, 1960, **74**(9, Whole No. 496).

Bolles, R. C., & Bailey, D. C. Importance of object recognition in size constancy. *J. exp. Psychol.*, 1956, **51,** 222–225.

Boring, E. G. *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century, 1942. Cited in J. S. Bruner, Jacqueline Goodnow, & G. A. Austin, *A study of thinking*. New York: Wiley, 1956.

Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64,** 123–152.

Bruner, J. S., Goodnow, Jacqueline, & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

Brunswick, E. *Perception and the representative design of experiments*. Berkeley: Univer. California Press, 1956.

Cronback, L. J., & Gleser, Goldine. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50,** 456–473.

Edwards, W. Behavioral decision theory. *Annu. Rev. Psychol.*, 1961, **12,** 473–498.

Ekman, G., Goude, G., & Waern, Y. Subjective similarity in two perceptual continua. *J. exp. Psychol.*, 1961, **61,** 222–227.

Goodnow, R. E. Utilization of partially valid cues in perceptual identification. Unpublished doctoral dissertation, Harvard University, 1954.

Irwin, F. W. Stated expectations as functions of probability and desirability of outcomes. *J. Pers.*, 1953, **21,** 329–333.

Irwin, F. W., Smith, W. A. S., & Mayfield, Jane F. Tests of two theories of decision in an "expanded judgment" situation. *J. exp. Psychol.*, 1956, **51,** 261–265.

Kelly, G. A. *The psychology of personal constructs*. Vol. 1. New York: Norton Press, 1955.

King-Ellison, Patricia, & Jenkins, J. J. The durational threshold of visual recognition as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67,** 700–703.

Nobel, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, **47,** 13–16.

Osgood, C. E., & Suci, G. J. A measure of relationship determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49,** 251–262.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana, Ill.: Univer. Illinois Press, 1957.

Overall, J. E., & Williams, C. M. Models for medical diagnosis. *Behav. Sci.*, 1961, **6,** 134–141.

Pollack, I., & Decker, L. R. Confidence ratings, message reception, and the receiver operating characteristic. *J. Acoust. Soc. Amer.*, 1958, **30,** 286–292.

Postman, L., & Rosenzweig, M. R. Practice and transfer in the visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 1956, **69,** 209–226.

Sarbin, T. R., Taft, R., & Bailey, D. E. *Clinical inference and cognitive theory.* New York: Holt-Rinehart, 1960.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43,** 195-201.

Summers, S. A. Learning of responses to multiple weighted cues. *J. exp. Psychol.*, 1962, **64,** 29-34.

Tanner, W. P., & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61,** 401-409.

(Received September 15, 1963)

Vol. 78, No. 8 Whole No. 585, 1964

Psychological Monographs: General and Applied

CONFLICT AND STRESS AS RELATED TO PHYSIOLOGICAL ACTIVATION AND SENSORY, PERCEPTUAL, AND COGNITIVE FUNCTIONING[1]

WALTER D. FENZ

University of Massachusetts

The study investigated the relationship between stress experienced in a real-life approach-avoidance conflict situation and physiological activation and performance. A 3-dimensional model permitted evaluation of conflict in relation to both temporal and cue dimensions. 27 sport parachutists were tested 3 different times in relation to a jump. The test consisted of stimuli scaled along a dimension of increasing relevance to parachuting. GSR and absolute level of conductance measured physiological activation. Level of performance was sampled over a wide range of responses, varying from simple sensory functions at one extreme to complex cognitive processes at the other.

In the present study novice parachutists were used as subjects (*Ss*) not only to investigate conflict in its own right, but also to explore the more general relationship between intensity of stress, on the one hand, and degree of physiological activation and performance deficit, on the other. Performance was sampled over a wide range of responses, varying from simple sensory functions at one extreme to complex cognitive processes at the other.

A difficulty that arises in investigating the effects of anxiety and conflict on performance is that the external criterion on which the selection of *Ss* is based is often no better than the dependent variable that is being evaluated. In the laboratory, on the other hand, where the variables under investigation are under adequate control, the conflict and anxiety that can be elicited are often not sufficiently intense and ego involving to produce meaningful results. A solution is provided by selecting certain real-life situations which can be manipulated in a controlled manner and involve intense levels of conflict. Such an approach retains the advantage of laboratory experimentation and, at the same time, involves an intensity of emotion that rivals and even exceeds the intensity of emotion observed in the behavior disorders.

A number of years ago, Luria (1932) measured behavior disorganization of students facing an important examination and of apprehended criminals facing trial. His results showed a breakdown of normal functioning upon presentation of relevant stimuli. In a more recent series of studies, using an approach similar to Luria's (Epstein & Fenz, 1962; Fenz & Epstein, 1962), the reactions of novice and experienced parachutists were investigated at different points in time in relation to an approaching parachute jump. Hypotheses were formulated within a framework of an approach-

---

[1] The present research was conducted in partial fulfillment of the requirements for the PhD degree under the supervision of S. Epstein at the University of Massachusetts. Appreciation is also extended to C. C. Neet and E. Dzendolet for their assistance. W. Main gratuitously offered office space at the Mansfield Municipal Airport. E. Getchell helped in the parachute training of *Ss*.

This investigation was supported by Research Grant M 1293 from the National Institute of Mental Health, United States Public Health Service, for which S. Epstein is principal investigator.

No. 7 wanting

avoidance conflict theory. The parachute jump was noted to be extremely ego involving and capable of eliciting intense affect. Additional advantages were that *Ss* could be tested in the laboratory, and level of conflict varied by manipulating the frequency and timing of the jump in relation to the testing session.

The experience of a first jump remains one to be talked about for many months to come. Only the most hardy continue the sport and become proficient parachutists. Follow-up statistics gathered in a large parachuting center (Istel, 1961) indicated that out of 2,800 first-time jumpers less than 15% made a second jump, and only a small fraction of these continued beyond the first steps of training. In a candid moment a beginning parachutist confided to the writer:

> It was one of the most exciting experiences in my life, but I sure would never, never ever want to go through it again.

There is little question that parachuting involves an intense state of stress and conflict for the novice jumper.

Model for the Measurement of Conflict

In 1956, Epstein presented a theoretical model of approach-avoidance conflict as applied to responses to projective techniques. His theoretical position is a synthesis of Miller's (1948, 1951) models of conflict and displacement with the psychoanalytic theory of thinking (Rapaport, 1951). Following the psychoanalytic model, Epstein assumed that with every drive state there is an associated tendency toward drive-related imagery (primary process) and also an image-inhibiting process. Drive-related expression is assumed to be analogous to approach, and drive-related inhibition to avoidance. Following Miller (1948), the assumption is made that the gradient of inhibition as a function of drive-relevant cues is normally steeper than the gradient of expression. Thus, verbal expression and inhibition are substituted for approach and avoidance, and a dimension of similarity of cues to the goal object is substituted for spatial distance.

Central to Epstein's (1956, 1962) formulation is the concept of *drive*, defined as a force with directive and energizing components. Conflict is viewed as an interaction of opposing drives. In this respect, Epstein's position is closer to Lewin's (1931) than to Miller's, whose gradients of approach and of avoidance represent responses or motor response tendencies. Epstein's concept of drive has the theoretical advantage of postulating a single concept from which two kinds of predictions and measurements can be derived: one relating to the activating component of drive, and the other to its directive component. Activation is considered a basic concomitant of all emotional states and is measured on an intensive dimension of physiological reactivity. Directionality relates to the qualitative distinction between approach and avoidance and can be inferred from content of verbal responses, i.e., the degree to which thoughts and their verbal expression move toward or away from a given area.

According to Epstein, conflict can be measured by three general indexes: one based on the net directive component of the two drives, one on the activating component, and one on both. More specifically, it is assumed that conflict is indicated by any of the following responses to a stimulus dimension of goal-relevant cues: (*a*) a relative increase in strength of approach responses to stimuli of low relevance, and a relative decrease in strength of approach responses to stimuli of high relevance; (*b*) a sharp rise in activation as a function of increasing stimulus relevance; (*c*) a decrease in adequacy of performance as a function of increasing stimulus relevance. This last hypothesis follows both from the assumption that high levels of activation are cognitively disruptive, and that the directive effects of strong expressive and inhibitory drives produce inappropriate overemphasis upon certain stimuli and avoidance of others. The qualification that the conflict must be of sufficient magnitude is added as it is assumed that the curve of adequacy of performance as a function of activation is inverted U shaped (Malmo, 1959).

Epstein and Fenz (1962) tested the

model's major hypotheses by investigating responses of sport parachutists to stimuli of various degrees of relevance to parachuting at various times from a parachute jump. Their findings indicated: (*a*) selective approach and avoidance to parachute-related words as revealed by perceptual sensitization and defense or by content of association; (*b*) increasing physiological activation along a dimension of increasing relevance to parachuting; (*c*) increasing performance deficit along the same stimulus dimension, as indicated by formal (noncontent) measures, such as reaction time. These findings confirmed the hypotheses on activation and response deficit; but the results on content, while significant, were not as predicted: with increasing proximity to a jump the content of associations showed an increasing focus of attention on the area of conflict and a relative insensitivity to other areas.

Statement of the Problem

The present study represents a replication and refinement of previous work on conflict over parachuting and includes a wider range of measures of the dependent variables.

The experiment derived from a three-dimensional model of conflict similar to those described by Murray and Berkun (1955) and by Epstein and Fenz (1962) in which conflict is considered in relation to a temporal and a cue dimension. The temporal dimension was manipulated by conducting a test on three occasions: 2 weeks before a jump, the day before a jump, and the morning of a jump. The cue dimension was created by varying the relevance to parachuting of stimuli in a word-association test. The study extended the time dimension from the two points investigated in earlier studies (Epstein & Fenz, 1962; Fenz & Epstein, 1962) to three points and substituted a scaled dimension of stimulus relevance for the previous ordinal scale, thereby permitting a more detailed description of the curves than simply noting whether they are monotonic gradients. Three response characteristics, presumed to be associated with conflict, were investigated: (*a*) activation; (*b*) performance deficit; and (*c*) directionality, i.e., approach and avoidance.

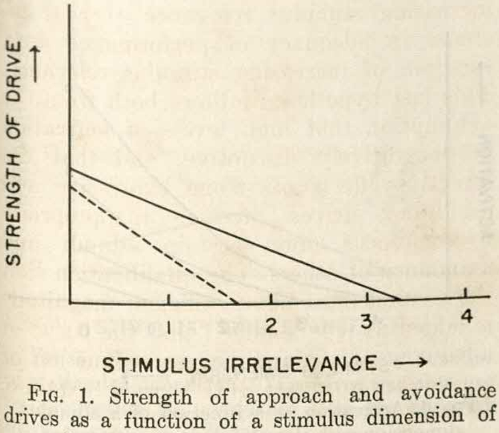


FIG. 1. Strength of approach and avoidance drives as a function of a stimulus dimension of cues relevant to parachuting.

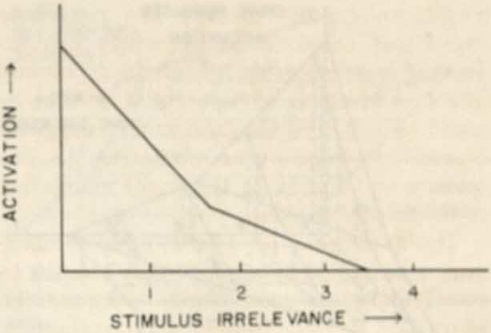


FIG. 2. Activation as a function of a stimulus dimension of cues relevant to parachuting. (This curve was obtained by adding the magnitudes of the approach and avoidance drives, presented in Figure 1.)

The model, adapted from Miller (1948), is presented in Figure 1. It represents approach as stronger at the goal than avoidance, since *S*s, by their own choice, advance to the goal. The model, so far, is two dimensional, as time is held constant. With regard to activation it is assumed that the magnitude of conflict-produced activation can be represented by the sum of the magnitudes of the approach and avoidance drives, disregarding algebraic sign (see Figure 2). The three-dimensional model of conflict, as applied to activation, is taken from an earlier study (Epstein & Fenz, 1962) and is presented in Figure 3. It will be noted that activation as a function of a time dimension is represented on the x axis,

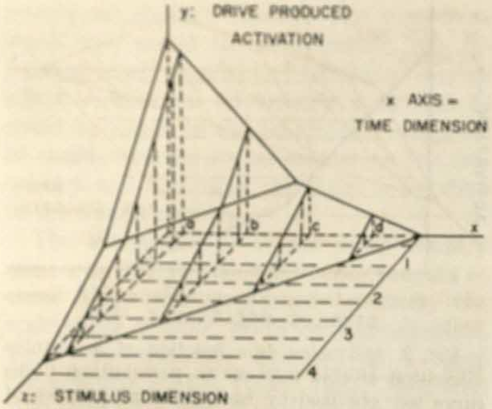


FIG. 3. Activation as a function of a stimulus and time dimension.

and a similar curve of activation as a function of a stimulus dimension is represented on the z axis. Four points are arbitrarily selected along the time dimension. The relationship of level of activation to the stimulus dimension at these four points in time is represented on the surface of the figure. The four resulting curves are extracted and presented as a family of curves in Figure 4. It is apparent that the gradients become higher and steeper as the time for the jump approaches. In a previous study (Epstein & Fenz, 1962), it was found that parachutists produced positive gradients of GSR as a function of a parachute-relevant stimulus dimension and that the gradients were steeper on the day of a jump than on a control day, consistent with the hypothesis. Control *S*s produced no gradients. The results were markedly reliable, remaining constant for every single *S*.

Turning now to performance deficit, it is assumed that degree of deficit is a direct function of activation, at least when the latter is within a range of high magnitude. Thus the curves of deficit as a function of cues and time should follow the theoretical curves for activation. In a previous study (Epstein & Fenz, 1962), response latency was used as a measure of performance deficit and gave results which were closely parallel to those for autonomic reactivity, as measured by GSR. The present study will investigate several measures of performance varying widely in degree of complexity and relevance to the area of conflict. This aspect of the study will yield information of general significance on the relationship between performance and activation, apart from its value for the measurement of conflict.

With regard to directionality of response, it is assumed that the magnitude of the drive-produced increment in net approach can be represented by the algebraic sum of the approach and avoidance drives (see Figure 5). The three-dimensional model of net approach as a function of time and cues (Epstein & Fenz, 1962; Fenz & Epstein, 1962) is presented in Figure 6. The curve for net approach increment as a function of the time dimension appears on the x axis, and a similar curve for net approach increment as a function of the stimulus dimension is presented on the y axis. The figure produced differs from the one for activation in that the addition of the avoidance to the approach gradient results in a change in the direction of the slope rather than simply a change in its steepness. Four planes parallel to the yz plane represent the same four time intervals as in the figure for activation (see Figure 3). The intersection of these planes with the surface of the figure represents the relationship of the net approach increment to the stimulus dimension for the time intervals in question. The curves are extracted from Figure 6 and presented as a family of curves in Figure 7. On the basis of this model, it had been predicted in earlier work (Epstein & Fenz, 1962;

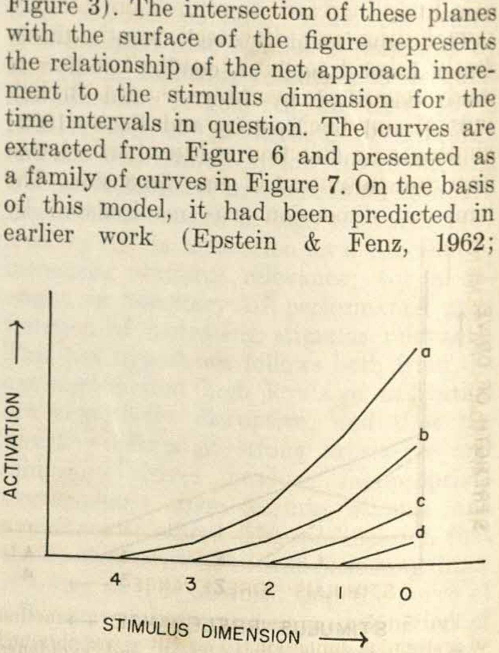


FIG. 4. Activation as a function of a stimulus dimension with time treated as a parameter.

Fenz & Epstein, 1962) that parachutists on the day of an anticipated jump produce stronger approach responses on a projective test to stimuli of low parachute relevance and weaker responses to stimuli of high parachute relevance than on a day 2 weeks from a jump. Contrary to prediction, however, an increase in strength of approach was noted for responses to both levels of stimuli, with the increase in response to stimuli of high relevance being the greater. These results would seem to support the assumption that the gradient of approach is steeper than the gradient of avoidance, rather than the reverse. The apparent discrepancy was resolved with the realization that projective tests measure expression and inhibition of verbal tendencies, rather than approach and avoidance in time or space. Such an expression-inhibition conflict in regard to the expression of fear of parachuting arises from the adaptive need to control and inhibit fear responses. On the assumption that the gradient of inhibition of verbal expression of fear is steeper than the gradient of its expression, results were consistent with the hypothesis, i.e., fear responses on the day of a jump increased to stimuli unrelated to parachuting and decreased to stimuli directly related to parachuting. The present study separately evaluates changes in feelings of approach and avoidance to parachuting as a function of time, and expression and inhibition of fear on a projective test as a function of cues in an attempt to demonstrate their functional independence.

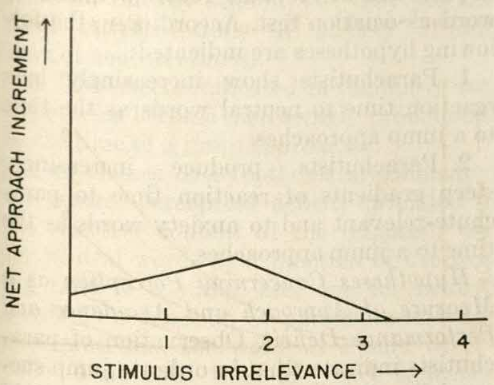


FIG. 5. Net approach increment as a function of a stimulus dimension. (This curve was obtained by subtracting the avoidance from the approach drive, represented in Figure 1.)

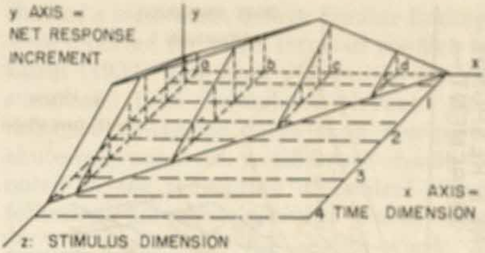


FIG. 6. Conflict produced net approach increment as a function of a stimulus and time dimension.

Formulation of Hypotheses

The following hypotheses stem from both the theoretical model described above, and previous empirical findings (Epstein & Fenz, 1962; Fenz & Epstein, 1962).

Hypotheses Concerning GSR as a Measure of Reactivity.

1. Parachutists become increasingly reactive physiologically, as indicated by GSRs to neutral words, as the time to a jump approaches.

2. Parachutists produce increasingly steep gradients of GSR to parachute-relevant and anxiety words as time to a jump approaches.

Both of these hypotheses follow directly from the model, and were substantiated in earlier work (Epstein & Fenz, 1962).

Hypotheses Concerning Absolute Level of Skin Conductance as a Measure of Activation. It is assumed that basal, or absolute, level of skin conductance varies directly with the emotional state of the organism. Unlike GSR, basal conductance does not involve an immediate response to stimulation (Woodworth & Schlosberg, 1954), but changes more gradually over time. It therefore does not provide a measure of the immediate impact of cues in a stimulus dimension. Nevertheless, since a measure of basal conductance was available, it was decided to investigate it, as this variable has too often been ignored (Woodworth & Schlosberg, 1954).

On the assumption that basal conductance is a measure of physiological activation, it should be relatively high prior to

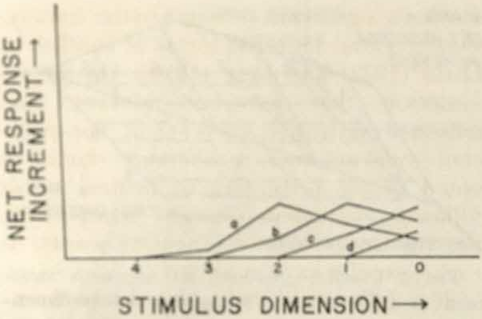


FIG. 7. Net approach increment as a function of a stimulus dimension with time treated as a parameter.

testing and should further rise while *S* is worrying about the nature of the test. After testing begins absolute conductance will rise or fall, depending upon whether the test is anxiety provoking, or upon whether it is not, and adaptation occurs. Because of the nature of the items, the test is assumed to be stressful for parachutists, but not for controls. It is further assumed that testing is more stressful for parachutists on the day of a jump than at a more remote time, as the parachutist is under greater conflict, and the completion of the test brings him one step closer to the actual jump. The following hypotheses are indicated:

1. Parachutists become increasingly reactive physiologically, as indicated by a rise in basal conductance, as the time to a jump approaches.

2. Parachutists and controls demonstrate a rise in absolute conductance during a 3-minute waiting period prior to testing.

3. During the testing itself, parachutists demonstrate a rise in absolute conductance level, with the sharpest rise occuring on the day of a jump.

4. Control *Ss* demonstrate a decline in absolute conductance during all testing sessions.

Hypotheses Concerning Auditory Threshold as a Measure of Performance Deficit. Changes in activation produced both by proximity to a jump and cues relevant to parachuting are assumed to raise auditory threshold. In the earlier study (Epstein & Fenz. 1962), it was found that parachutists on the day of a jump exhibited perceptual deficit for neutral and anxiety words. The question may be raised as to whether the deficit, particularly for neutral words, was directional and due to focusing on parachute-relevant cues, or whether a more general sensory deficit was involved. It was decided to test this further by the use of a measure of auditory threshold, which is relatively independent of the directional elements of the approach and avoidance drives, since the responses, involving recognition of pure tones, cannot be classified according to relevance to parachuting.

The following hypotheses are indicated:

1. Parachutists show increasing auditory deficit for tones presented after neutral words, as the time to a jump approaches.

2. Parachutists produce increasingly steep gradients of auditory deficit to pure tones as a function of the parachuting relevance of preceding cues, as the time to a jump approaches.

Hypotheses Concerning Response Latency as a Measure of Performance Deficit. The question may be raised as to whether response latency should be considered a measure of deficit or of directionality. In animal studies (Kimble, 1961) reaction time is frequently used as a measure of approach and avoidance. However, it is also known that strong states of tension may disrupt thinking and produce blocking (Luria, 1932). Previous findings (Epstein & Fenz, 1962) support the latter interpretation of elevated reaction time in a word-association test. Accordingly, the following hypotheses are indicated:

1. Parachutists show increasingly long reaction time to neutral words as the time to a jump approaches.

2. Parachutists produce increasingly steep gradients of reaction time to parachute-relevant and to anxiety words as the time to a jump approaches.

Hypotheses Concerning Perception as a Measure of Approach and Avoidance and Performance Deficit. Observation of parachutists indicates that in order to jump successfully they must control their fears. At some parachuting centers it is common practice to sing forceful parachuting songs during ascent in the plane. It is assumed that this does not allow the beginning jumper time for frightening thoughts. There is often jovial conviviality and much reassuring patting on the helmet and shoulders, which makes the beginning jumper feel that the jumpmaster is right with him and that he has nothing to fear. To the extent that parachutists are attempting to emphasize their approach reactions to parachuting, and de-emphasizing their fear reactions, they should demonstrate perceptual sensitization to parachute-relevant stimuli and perceptual defense to anxiety-producing stimuli. Earlier findings (Epstein & Fenz, 1962) were consistent with this analysis, demonstrating the appropriate selective effects which were superimposed upon a more general perceptual deficit associated with an approaching jump. Thus, selective perception provides an index of approach and avoidance and can be used to measure degree of approach to parachute-relevant cues and of avoidance of anxiety cues. In addition, nonselective deficit can be measured by perception of stimuli unrelated to parachuting.

The following hypotheses are indicated for the degree of approach and avoidance manifested in the perceptual response:

1. Parachutists exhibit perceptual sensitization for parachute-relevant stimuli, i.e., they produce fewer errors of perception in responding to parachute-relevant stimuli than in responding to unrelated stimuli.

2. Parachutists exhibit perceptual defense for anxiety stimuli, i.e., they produce more misperceptions of anxiety stimuli than of neutral stimuli.

3. The effects indicated in the above two hypotheses become increasingly pronounced as the time to a jump approaches.

Two classes of hypotheses are indicated for general perceptual deficit, unrelated to the selective content of the response: one for neutral words as a function of decreasing time to a jump, and one for neutral words as a function of parachute-relevant cues preceding the word. It should be recalled that both time and cues influence activation and thereby performance. It was found in an earlier study (Epstein & Fenz, 1962) that perception of neutral words *following* parachute-relevant words showed a significant deficit. Similar findings were reported for other forms of conflict by Luria (1932) and Jung (1919). On the assumption that parachute-relevant cues influence activation more than non-parachute-relevant cues, a difference should be noted in the perception of neutral words following parachute-relevant words from neutral words following neutral words, as earlier studies have indicated.

The following hypotheses for performance deficit in perception are thus indicated:

1. Parachutists demonstrate increasing misperceptions of neutral stimuli as time to a jump approaches.

2. Parachutists produce more misperceptions of neutral words following parachute-relevant words than of neutral words following neutral words.

3. The effect indicated in the preceding hypothesis becomes increasingly pronounced as time to a jump approaches.

Hypotheses Concerning Memory as a Measure of Approach and Avoidance and Performance Deficit. Memory, like perception, is assumed to reveal both selective effects and the effect of general deficit. The effects of sensitization (approach) can be demonstrated by selective recall of parachute-relevant material, of defense (avoidance) by selective forgetting of anxiety-arousing material, and of general deficit by a failure to recall neutral stimuli.

Following are the hypotheses for directionality and general deficit in regard to memory.

Directionality.

1. Parachutists recall more parachute-relevant stimuli than neutral stimuli relative to control *S*s.

2. Parachutists recall fewer anxiety-arousing stimuli than neutral stimuli relative to control *S*s.

3. The effects indicated in the preceding two hypotheses become increasingly pronounced as time to a jump approaches.

Performance deficit. Parachutists show increasing memory deficit for neutral stimuli as time to a jump approaches.

Hypotheses Concerning Content of Association as a Measure of Approach and Avoidance. Content of association is assumed to reflect the adaptive needs of the parachutists in coping with cues relating to parachuting and fear of parachuting. In an earlier study (Epstein & Fenz, 1962), it was found that it was difficult to evaluate reliably the relevance to parachuting of verbal responses of *S*s. Accordingly, all parachute-relevant responses were combined into one overall score. On the basis of findings in the earlier study, and consistent with the theoretical analysis of expression and inhibition of approach and avoidance in projective tests, the following hypotheses are indicated:

1. Parachutists produce more parachute-relevant associations than control *S*s.
2. Parachutists produce an increasing number of parachute-relevant associations as time to a jump approaches.
3. Parachutists produce fewer anxiety responses as time to a jump approaches.

Hypothesis Concerning Self-Ratings as a Measure of Approach and Avoidance. In order to test observations that novice parachutists seem highly motivated to jump at a time remote from a jump, but not at all eager when getting strapped into the harness and during ascent in the plane, a formal test was made requiring *S*s to rate their feelings of approach and avoidance along a sequence of events before and after a jump. Frequently a beginning jumper reports how, after having arrived at the airport, and while getting ready to make his jump, and especially during ascent in the plane, he wondered why he ever got himself into this predicament, and how he could possibly get out of it without losing face. On the assumption that the gradient of avoidance is steeper than the gradient of approach, the following hypothesis for self-ratings of approach and avoidance is indicated: Parachutists report a decrease in feelings of approach and an increase in feelings of avoidance as time to a jump approaches, so that the degree of dominance of approach over avoidance steadily diminishes.

Hypotheses Concerning Thematic Responses as a Measure of Approach and Avoidance. Earlier studies on parachuting (Epstein & Fenz, 1962; Fenz & Epstein, 1962) indicated that parachutists express rather than inhibit parachute-relevant thoughts, despite their anxiety-arousing potential. Such a reaction tendency is obviously adaptive, as the parachutist must be prepared to deal with such cues. It was found that with increasing proximity to a jump, in addition to an increase in expression of approach responses to jumping, there was an increase in defensive operations involving fear, such as displacement and denial. In this respect, *S*s were inclined to express fear to stimuli unrelated to parachuting and to deny fear to stimuli strongly related to parachuting. The present hypotheses for the directional content of the thematic responses are derived from these findings. Since in the present study a thematic apperception test is administered only once, no comparisons on time effects can be made; and the following hypotheses relate only to the differential effect of the stimulus dimension:

1. Parachutists obtain higher scores of approach to parachute-relevant stimuli at all levels of relevance than control *S*s. However, the difference is greater for stimuli of higher relevance to parachuting than for stimuli of moderate relevance.
2. Parachutists obtain lower scores of fear on stimuli highly related to parachuting and higher scores of fear on stimuli unrelated to parachuting than control *S*s.

METHOD

The Ss

Twenty-seven college students who had expressed an interest in sport parachuting constituted the experimental group in this study. They were recruited at parachuting centers in the Boston area and were remunerated for the cost of training up to, and including, their second parachute jump, in exchange for taking part as *S*s in the experiment. All experimental *S*s at the time of testing had been instructed in the rudiments of parachuting, had made one or two jumps prior to testing, and had agreed to make at least one more jump as part of the research program. Parachutists were not tested prior to their first jump, because in the earlier studies on parachuting it had been observed that the experience of having made a first jump increased apprehension at the next jump. The experimental group was matched with a comparable group of 27 control *S*s selected from students in an introductory course in psychology at the University of Massachusetts who volunteered to take part in experiments. The age of experimental and control *S*s in the experiment ranged from 18 to 23, and all *S*s were enrolled as full-time college or university students at the time of testing.

Research Instruments

Word-Association Test. It was necessary to select words along a dimension of increasing relevance to a critical event (the parachute jump) to be used as stimuli in the word-association test. Since the test was to be administered to both experimental and control *S*s, it was important that both groups of *S*s perceive the stimulus words as equally relevant to parachuting, or else any differences between the two groups could be attributed to this factor alone. Furthermore, as all *S*s were to be tested three times, it was necessary to prepare three parallel forms of the word-association test.

The method used in scaling the stimulus dimension had the advantage of utilizing the judgments of the experimental groups themselves, thereby assuring that the experimental and control groups perceived the stimuli in the same manner. Finally a scale of quasi-interval type could be obtained, which allowed for a more refined statistical analysis than would otherwise be possible.

The procedure used in scaling can be classified among the methods of subjective estimate, in which *S*s are instructed to estimate quantitatively the value of a number of stimuli (Woodworth & Schlosberg, 1954). The subjective estimates were scaled by the method of successive intervals, as described by Edwards (1957). Figure 8 presents the empirical means obtained through the judgment method, the scale values for the four classes of stimuli, and the boundaries of the respective stimulus categories. An internal consistency check, described by Edwards (1957), yields an average error of .015, which is fairly typical of the values reported for the average error obtained when this method of attitude scaling is used. Edwards and Thurstone (1952), for example, report an average error of .025 for 10 stimuli scaled into 9 categories; and Edwards (1952) reports a value of .021 for 17 stimuli scaled into 10 categories.

Each word list contains three words per stimulus category, that is, three neutral, three low, three medium, and three high relevant words, and, in addition, three anxiety words. The words are arranged randomly and are interspersed among nonscaled "buffer" words. Five practice words occur at the beginning of the list.

Following are the word lists with the scaled stimuli designated by the letter N for neutral, L for low, M for medium, H for high, and A for anxiety.

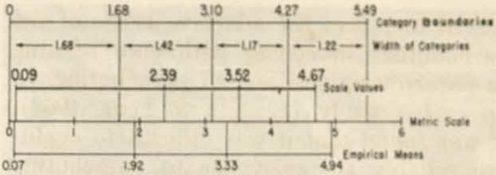


FIG. 8. Empirical means, estimated scale values, and category boundaries of stimuli rated for relevance to parachuting.

1. Word List A: FUN, DARK, WISH, SMOOTH, SPIDER, **OPENED**-M, CASTLE, **SKY**-L, WORD, **KILLED**-A, MOON, GIRL, **MUSIC**-N, SPOON, **HURT**-A, PINK, **PARACHUTE**-H, SUN, **SWIFT**-L, WOMAN, OCEAN, **STOVE**-N, LION, **PAPER**-N, HIPPOPOTAMUS, **AIRCRAFT**-M, HORSE, **AIRPORT**-L, BASEBALL, **RIPCORD**-H, FIREPLACE, BOOK, **ANXIETY**-A, HEAVY, **FLYING**-M, FACE, **FALL**-H.

2. Word List B: JOY, CHEESE, FUN, SCHOOL, SLOW, **HEIGHT**-M, RIVER, **COVERALLS**-L, TABLE, **FEAR**-A, LIGHT, PEN, **SALT**-N, DAY, **INJURY**-A, CHICKEN, **SKYDIVER**-H, HUNGRY, **TAXI**-L, RADIATOR, HAPPY, **BICYCLE**-N, CARPET, **BLACK**-N, GREEN, **ALTITUDE**-M, BEAUTIFUL, **TAKE-OFF**-L, CHILD, **BAILOUT**-H, FOOTBALL, MAN, **BLACKOUT**-A, NUMBER, **PILOT**-M, RHINOCEROS, **JUMPMASTER**-H.

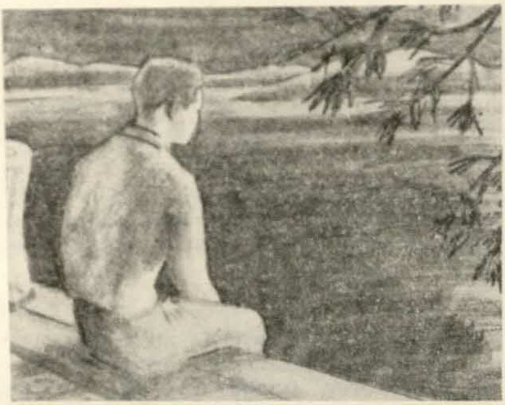
3. Word List C: WALL, RED, HAPPY, DESK, FOX, **AIRPLANE**-M, UNCLE, **WINGS**-L, RIDE, **ACCIDENT**-A, MEMORY, THIRST, **CHAIR**-N, SEA, **FATALITY**-A, TIGER, **PARATROOPER**-H, ELEPHANT, **RESERVE**-L, READ, TENNIS, **LAMP**-N, KING, **PENCIL**-N, CAB, **STREAMER**-M, RADIATOR, **FAST**-L, WATCH, **JUMP**-H, SOFT, EAT, **PANIC**-A, TOWN, **ALTIMETER**-M, LETTER, **FREEFALL**-H.

Words were presented by a tape recorder at 30-second intervals after having first been screened for clarity of pronunciation. The *S*s were instructed to respond as quickly as possible with the first word that occurred to them. Response words were recorded by the examiner.

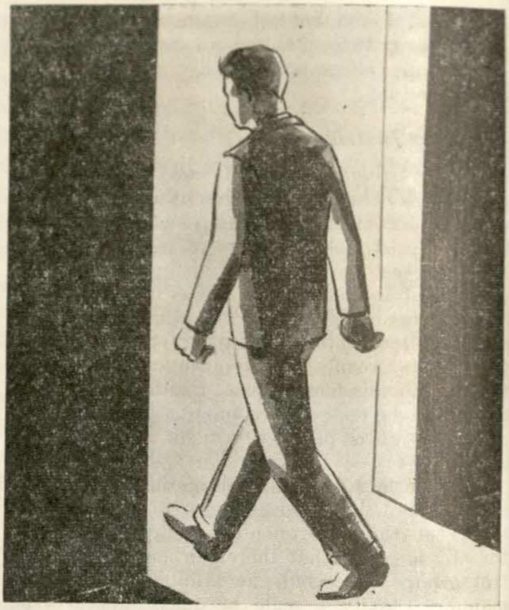
Thematic Apperception Test. This test consisted of pictures constructed by a professional artist: four unrelated to parachuting, one slightly related to parachuting, and one strongly related to parachuting. The pictures are presented in Figure 9. All of these pictures were previously used in a study by Fenz and Epstein (1962).

The six pictures were presented in the following order: neutral, neutral, slight relevance, neutral, high relevance, and neutral. The purpose of the first neutral picture was to allow for adaptation, while the other three neutral pictures served as buffers.

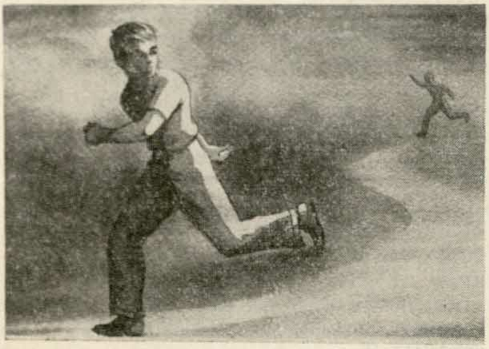
Parachutists were administered the set of pictures only on the day of the jump. The thematic test was presented after the word-association test, so that there was no possibility of contaminating the latter. Control *S*s were tested in sessions corresponding in time to the sessions for the parachutists. Responses of *S*s were tape recorded, transcribed, coded, and scored blindly.



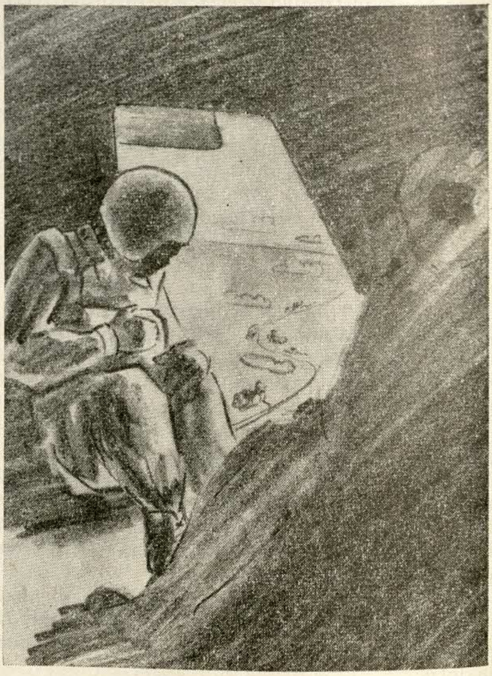
9a. Neutral.



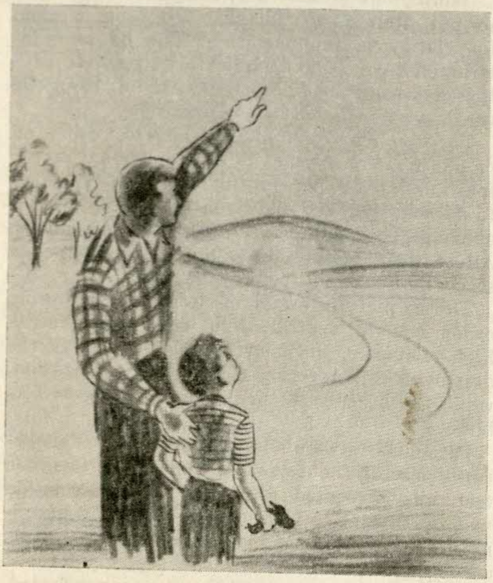
9d. Neutral.



9b. Neutral.

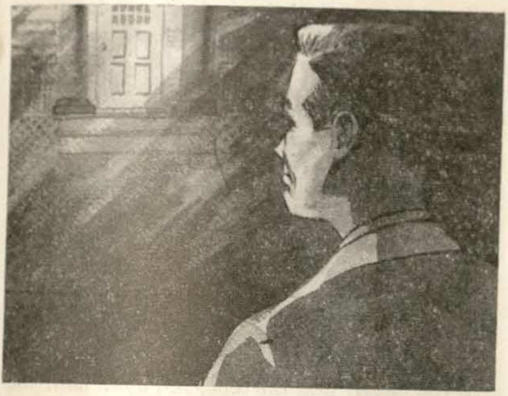


9e. High relevance.



9c. Slight relevance.

(Fig. 9 continued on the next page.)



9f. Neutral.

FIG. 9. Thematic picture stimuli.

Measures

The following dependent variables were investigated:

- I. GSR
  - A. Response to neutral words
  - B. Response to words on a dimension of relevance to parachuting
  - C. Response to anxiety words
- II. Absolute level of skin conductance
  - A. Level of conductance preceding each of the three testing sessions
  - B. Level of conductance during a 3-minute interval immediately prior to the beginning of the word-associaiton test
  - C. Level of conductance during course of testing
- III. Auditory threshold
  - A. Auditory threshold following stimulation by neutral words
  - B. Auditory threshold following stimulation by words on a dimension of relevance to parachuting
  - C. Auditory threshold following stimulation by anxiety words
- IV. Response latency
  - A. Response latency to neutral words
  - B. Response latency to words on a dimension of relevance to parachuting
  - C. Response latency to anxiety words
- V. Perception
  - A. Misperception of neutral words
  - B. Misperception of parachute-relevant words
  - C. Misperception of anxiety words
  - D. Misperception of neutral words following relevant words
- VI. Memory
  - A. Recall of neutral words
  - B. Recall of parachute-relevant words
  - C. Recall of anxiety words
- VII. Association
  - A. Parachute-relevant associations
  - B. Anxiety-relevant associations
- VIII. Subjective estimates
  - A. Approach estimated along a time dimension
  - B. Avoidance estimated along a time dimension
- IX. Thematic responses
  - A. Net approach content
  - B. Fear of parachuting
  - C. Displacement and denial of fear of parachuting

Apparatus

Recordings of skin resistance were obtained from a direct-current Hunter GSR apparatus, Model No. 300, operating on the principle of the Wheatstone bridge. The GSR response was recorded from the index and middle fingers of the nondominant hand, using nonpolarizing zinc electrodes and zinc sulphate paste. Resistance was registered by an automatic recording pen. Reaction time was recorded by a penmarker controlled by the examiner.

Absolute auditory thresholds were obtained for a tone of 800 cps. The output control of a General Radio beat-frequency oscillator (Type No. 1304-AO) was placed at its O-db. setting (reference level of 1 milliwatt into 600 ohms), and the output was led into a Hewlett-Packard attenuator (Model 350 B). An attenuator setting for each *S* was chosen so as to be near his binaural absolute threshold, under the environmental conditions available. The output was finally led into the earphones (Permoflux Corporation), worn by *S* and also into a rectifying circuit whose output led into an oscillograph. The *S* changed the loudness of the tone by moving the output control of the oscillator. The rectifying circuit allowed a record of *S*'s absolute threshold to be displayed on the oscillograph. The record was calibrated in decibels.

Office space was provided at the Mansfield Municipal Airport. Temperature was kept constant at 70° F.

Experimental Procedure

The choice of an experimental design required consideration of the following: (*a*) the effect of sequence, since all *S*s were tested three times; (*b*) the conditions of testing, that is, whether testing was on the day of a jump, the day before a jump, or 2 weeks before a jump; and (*c*) possible differences in the three word lists. These three variables were counterbalanced according to the design shown in Table 1.

All 27 parachutists were tested once on the day of a jump, once the day before a jump, and once 2 weeks before a jump. Of these, nine *S*s were tested first on the day of a jump, nine first on the day before a jump, and nine first on a day 2

TABLE 1

EXPERIMENTAL DESIGN

*S*s	Design		
1–3	$A_1B_1C_1$	$A_2B_2C_2$	$A_3B_3C_3$
4–6	$A_1B_1C_2$	$A_2B_2C_3$	$A_3B_3C_1$
7–9	$A_1B_1C_3$	$A_2B_2C_1$	$A_3B_3C_2$
10–12	$A_1B_3C_1$	$A_2B_1C_2$	$A_3B_2C_3$
13–15	$A_1B_3C_2$	$A_2B_1C_3$	$A_3B_2C_1$
16–18	$A_1B_3C_3$	$A_2B_1C_1$	$A_3B_2C_2$
19–21	$A_1B_2C_1$	$A_2B_3C_2$	$A_3B_1C_3$
22–24	$A_1B_2C_2$	$A_2B_3C_3$	$A_3B_1C_1$
25–27	$A_1B_2C_3$	$A_2B_3C_1$	$A_3B_1C_2$

Note.—A stands for condition of testing (A_1 = day of jump, A_2 = day before jump, and A_3 = 2 weeks from jump). B stands for sequence (B_1 = first testing, B_2 = second testing, and B_3 = third testing). C for word list used (C_1 = Word List A, C_2 = Word List B, and C_3 = Word List C).

weeks from a jump. Of the nine *S*s who were tested first on the day of a jump, three were first tested with Word List A, three with Word List B, and three with Word List C. Testing of control *S*s followed a parallel design, with the only difference that condition of testing had no specific meaning, except in regard to the time interval between sessions.

Upon entering the experimental room and being seated, a brief explanation was given about the word-association test, during which *S* was told that the purpose of the test was to measure emotion in relation to parachuting. Instructions included three points: the requirements of the word-association test, the method of continuously manipulating the intensity of a pure tone at threshold, and the information that *S* would be asked to recall the stimulus and response words at the end of the test.

Having put on the earphones, *S* was told to report "yes" when he heard a tone and to indicate "no" when he was no longer able to hear it. He was also told to disregard incidental audible noises such as might be emitted by the apparatus or the surrounding environment. The experimenter manipulated stimulus intensity prior to the actual experiment in order to determine *S*'s absolute threshold. During the word-association test, *S* continuously manipulated the intensity level of the tone by raising it from just above to just below threshold, and vice versa. The procedure can best be described by the actual instructions:

> You are to turn the knob clockwise, slowly, and in small fractions, up to the point when you clearly hear the sound; having reached that point, start turning the knob anticlockwise, down to when you no longer hear it. You are to continue this operation without interruption throughout the testing.

When the experimenter had assured himself that *S* was able to follow the instructions, he placed the GSR electrodes on the left-hand index and middle fingers. After 3 minutes, to permit *S* to adapt to them, the tape-recorded list of words was introduced with the following comment:

> Now we will begin a test of speed of association to words. After you have heard a word—and not before you have heard a whole word—say the first word that occurs to you as quickly as possible. If you are not sure of a word, respond to what you think it was. At the end of the test you will be asked to remember as many of the stimulus and the response words as you can. While responding to the word heard, you are to continue to slowly turn the knob up until you clearly hear the tone, and down until you no longer hear it, throughout the testing. Please do not make any comments or ask any questions between words, but save them for the end of the test. Do you have any questions now?

Scoring of Responses

Figure 10 represents a sample from the actual recording of responses of a parachutist on the day of a jump. Reaction time represents time from end of stimulus to beginning of response. Since the recording paper moved at a constant speed of 5 mm. per second, 8.5 mm. represents a reaction time of 1.70 seconds.

GSR was measured by the magnitude of the increase in conductance immediately following presentation of a stimulus word. A represents base level of conductance, B prestimulus, and C poststimulus levels. The reactivity measure was obtained by subtracting B from C.

The psychophysiological method used was a modified method of limits. All values of the ascending order were averaged, and all values of the descending order were averaged. The obtained averages in turn were averaged, giving the final threshold measure. The time limit within which changes were scored was from immediately after presentation of the stimulus to 15 seconds later, i.e., the half-way point between stimuli. This included about 3–5 ascending and descending series following each stimulus presentation. In Figure 10, Point A represents the zero point; the ascending order moves toward the center of the paper, and the descending order away from the center.

Scoring of measures involving verbal content will be described in the section on results.

Statistical Procedures

Where distributions permitted, the data were analyzed by analysis of variance. In one analysis experimental *S*s were used as their own controls, and nonparachutists were omitted. This permitted an orthogonal arrangement of experimental condition and sequence. In a parallel analysis, nonparachutists were used as their own controls and compared on first, second, and third testings. In a third analysis which included parachutists and

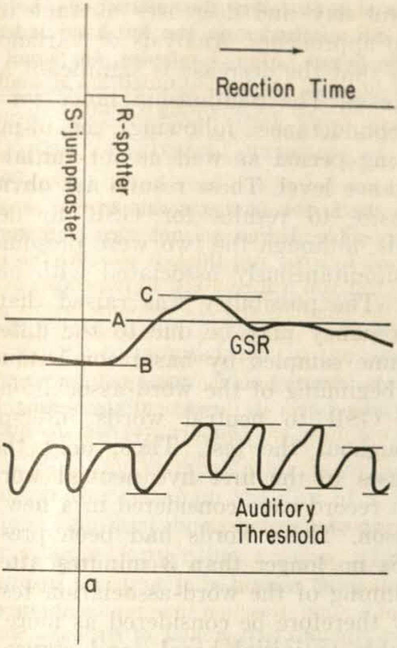


FIG. 10. Sample of galvanic skin response, reaction time, and auditory threshold recordings.

nonparachutists, it was necessary to sacrifice information on the interaction of sequence and experimental condition, as this was meaningless for nonparachutists.

RESULTS

The GSR

As had been hypothesized, it was found that GSR is directly related (*a*) to the relevance of the eliciting stimulus to parachuting and (*b*) to closeness in time to a jump. As may be seen in Figure 11, parachutists produce gradients of GSR, the steepness of which varies directly with proximity to a jump. Control *S*s fail to produce a gradient.

Three analyses were performed: one using parachutists as their own controls, one comparing control *S*s at the three testing sessions, and one comparing parachutists with control *S*s. These are shown in Tables 2, 3, and 4. The increase in steepness of gradients as a function of time to a jump is highly significant (.001 level), while the corresponding source of variance for control *S*s does not approach significance.

Trend analysis of the significant interaction demonstrated by the parachutists on the variables of experimental condition (time) and stimulus dimension (cues) indicates that the significance can be attributed to a combination of linear, quadratic, and cubic components. Individual analyses carried out for each experimental condition reveals increasing linear, quadratic, and cubic functions from control day to day of jump. As may be seen in Figure 11, the gradients, apart from becoming generally steeper, become increasingly positively accelerated as time to a jump approaches.

Turning back to Figure 11, it can be seen that parachutists react more strongly to neutral words on the day of a jump and the day before a jump than on a day 2 weeks from a jump. This difference, tested by analysis of variance, is significant at the .05 level. The difference between the day before and the day of a jump is negligible. It may be concluded that the level of reactivity of parachutists increases up to a point as the time for the anticipated jump approaches. Parachutists are also shown to react less strongly to neutral words on a control day than their nonparachuting counterparts. This difference is significant at the .01 level. Apparently, when not jumping, parachutists tend to react less strongly than nonparachutists to stimuli that are not associated with parachuting. This finding appears to reveal a personality difference as it may be interpreted to indicate that parachutists are more apt to release tension in motor activity, rather than to inhibit their emotions, or bind their tensions.

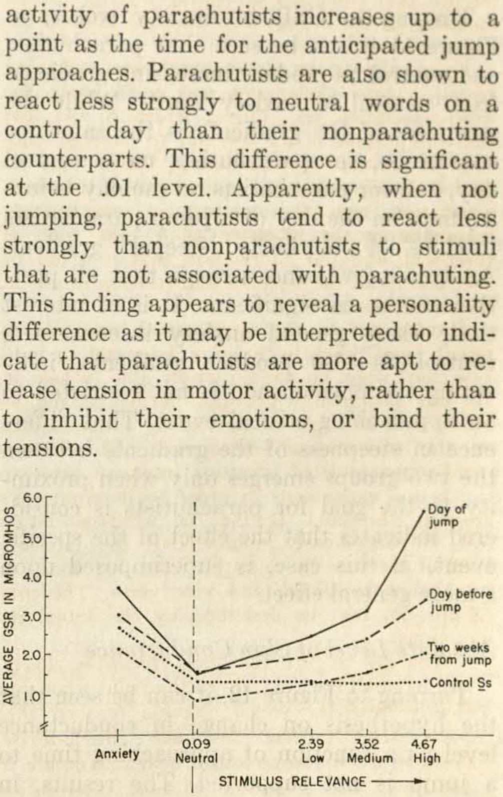


FIG. 11. GSR of parachutists and control *S*s to a stimulus dimension and anxiety words.

TABLE 2

ANALYSIS OF VARIANCE OF GSR TO A STIMULUS DIMENSION FOR PARACHUTISTS ON THE DAY OF A JUMP, THE DAY BEFORE A JUMP, AND ON A CONTROL DAY

Source	*df*	*MS*	*F*
Between *Ss*	26		
Sequence (Q)	2	24.75	2.58
Ss/Q	24	9.58[a]	28.18***
Within *Ss*	297		
Experimental Condition (E)	2	83.15	29.42***
Q × E	4	7.53	2.66*
Ss × E/Q	48	2.83[a]	8.32***
Dimension (D)	3	83.17	86.61***
Q × D	6	.82	.85
Ss × D/Q	72	.96[a]	2.64***
E × D	6	12.77	37.54***
Q × E × D	12	.19	.56
Ss × E × D/Q	144	.34[a]	

[a] Error term for *MSs* above it up to next error term. The *Ss* × E × D/Q interaction was used to evaluate all other error terms.
* $p < .05$.
*** $p < .001$.

Turning to GSR to anxiety words, in Figure 11 it can be seen that parachutists and control *Ss* produce increasing gradients from neutral to anxiety words. While the steepness of the gradients is the same for control *Ss*, for parachutists on a control day, and for parachutists on the day before a jump, on the day of a jump parachutists produce a markedly steeper gradient. Analysis of variance reveals that all positive effects are significant. It is of interest that, where general anxiety is concerned, control *Ss* also produce gradients, indicating the phenomenon is not restricted to an approaching critical event. That difference in steepness of the gradients between the two groups emerges only when proximity to the goal for parachutists is considered indicates that the effect of the specific event, in this case, is superimposed upon a more general effect.

Absolute Level of Skin Conductance

Turning to Figure 12, it can be seen that the hypothesis on change in conductance level as a function of approaching time to a jump is not supported. The results, in fact, are in the opposite direction of the hypothesis. Conductance is highest on the control day and decreases as time to the jump approaches. Analysis of variance reveals that the decrease is significant at the .05 level. The relationship holds for level of conductance following the 3-minute waiting period as well as for initial conductance level. These results are obviously opposite to results for GSR to neutral words, although the two were presumed to be simultaneously associated with activation. The possibility was raised that this discrepancy may be due to the difference in time sampled by basal conductance at the beginning of the word-association test, and GSR to neutral words interspersed throughout the test. Thus, only the responses to the first five neutral words of each record were considered in a new comparison. These words had been presented to *Ss* no longer than 3 minutes after the beginning of the word-association test and may therefore be considered as more comparable to initial basal conductance. The means for the five words at the beginning of the test were found to be somewhat higher than the means for the neutral words interspersed throughout the test, but do not differ in relative position, i.e., parachutists on the day 2 weeks before a jump obtain smaller GSRs than on the day be-

TABLE 3

ANALYSIS OF VARIANCE OF GSR TO A STIMULUS DIMENSION FOR CONTROL *Ss* AT THREE TESTING SESSIONS

Source	*df*	*MS*	*F*
Between *S*s	26		
Sequence (Q)	2	2.57	.32
*S*s/Q	24	8.03[a]	25.90***
Within *S*s	297		
Testing Sessions (T)	2	4.54	.88
T × Q	4	9.71	1.88
*S*s × T/Q	48	5.15[a]	16.61***
Dimension (D)	3	.35	1.59
D × Q	6	.12	.55
*S*s × D/Q	72	.22[a]	23.18***
T × D	6	.45	1.16
T × D × Q	12	.53	1.72*
*S*s × T × D/Q	144	.31[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × T × D/Q interaction was used to evaluate all other error terms.
* $p < .05$.
*** $p < .001$.

fore a jump or the day of a jump. Thus the discrepancy between the findings on GSR and basal conductance cannot be attributed to differences in the periods sampled, and it must be concluded that they are not functionally equivalent measures of activation.

How is one to account for this unexpected finding? One possibility is that a blanket type of inhibition of anxiety occurs as the time to a jump approaches, which succeeds in lowering the general resting level of conductance in the absence of strong stimulation. This general state of inhibition might then be disrupted by specific stimulation which demands a reaction. Such an interpretation is supported by the findings that on the day of a jump absolute conductance before testing is lower than at any other session, while at the end of the test it is higher than at any other session.

In Figure 12 it can further be seen that the other hypotheses involving level of conductance are generally verified. During the 3-minute period before testing, absolute level of conductance rises for all *S*s. Following this, control *S*s demonstrate adaptation by a drop in conductance level; while parachutists exhibit a sharp increase in conductance on the day of a jump, less of a rise on the day before a jump, and a slight decline 2 weeks from a jump.

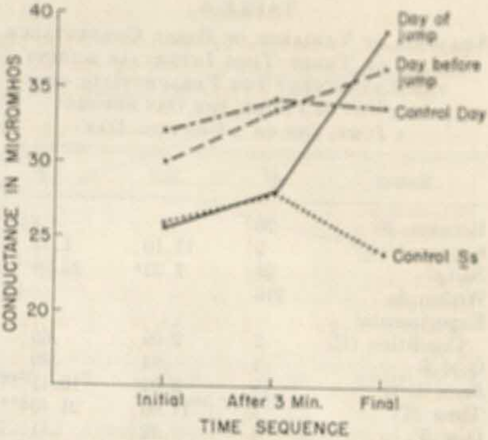


FIG. 12. Absolute level of skin conductance of parachutists and control *S*s as a function of testing sessions.

TABLE 4

ANALYSIS OF VARIANCE OF GSR TO A STIMULUS DIMENSION FOR PARACHUTISTS AND CONTROL *S*s

Source	*df*	*MS*	*F*
Between *S*s	53		
Group (G)	1	162.10	17.66***
*S*s/G	52	9.18[a]	27.82***
Within *S*s	594		
Experimental Condition (E)	2	24.43	5.62**
G × E	2	63.26	14.56***
*S*s × E/G	104	4.34[a]	13.15***
Dimension (D)	3	42.82	73.53***
D × G	3	40.71	69.91***
*S*s × D/G	156	.58[a]	1.75***
D × E	6	6.75	20.67***
G × D × E	6	6.37	19.53***
*S*s × E × D/G	312	.33[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × E × D/G interaction was used to evaluate all other error terms.

** $p < .01$.

*** $p < .001$.

In order to establish the reliability of the above observations, the usual analyses of variance were performed. When parachutists were tested as their own controls (see Table 5), the main effect of time (i.e., the three periods in which the testing session is divided), and the interaction of time period and experimental condition were highly significant. The significant main effect is due to the overall rise in level of conductance of parachutists throughout the testing sessions, when all three sessions were averaged; and the significant interaction is due to the sharper rise in level of conductance on the day of a jump than at either of the other times. The analysis for control *S*s (see Table 6) fails to reveal any source of significance other than individual differences, indicating that the adaptation effect that appears in Figure 12 is not reliable. When parachutists and controls are compared in a single analysis (see Table 7), the groups differ significantly in their pattern of conductance change as a function of the time period within the testing session, confirming the results obtained when the groups are analyzed independently.

Figure 13 presents the average prestimulus levels of basal conductance at four

TABLE 5

ANALYSIS OF VARIANCE OF BASAL CONDUCTANCE TAKEN AT THREE TIME INTERVALS DURING THE EXPERIMENT FOR PARACHUTISTS ON THE DAY OF A JUMP, THE DAY BEFORE A JUMP, AND ON A CONTROL DAY

Source	*df*	*MS*	*F*
Between *S*s	26		
Sequence (Q)	2	11.10	1.53
*S*s/Q	24	7.22[a]	24.98***
Within *S*s	216		
Experimental Condition (E)	2	2.09	.69
Q × E	4	.93	.30
*S*s × E/Q	48	3.02[a]	10.41***
Time (T)	2	11.00	21.40***
Q × T	4	.22	.41
*S*s × T/Q	48	.52[a]	1.75
E × T	4	3.29	11.36***
Q × E × T	8	.21	.72
*S*s × E × T/Q	96	.29[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × E × T/Q interaction was used to evaluate all other error terms.

*** $p < .001$.

intervals during the word-association test. No analysis was performed on these data, which are presented only to illustrate in a more detailed manner the effects that appear in Figure 11 for the last two intervals. It is evident that for parachutists on the day of a jump, the increase in conductance is continuous throughout the testing session.

Auditory Threshold

Results on auditory threshold are shown in Figure 14. As had been hypothesized, findings reveal an increase in steepness of the threshold gradient with increasing proximity to a jump. On the day of a jump, auditory thresholds obtained after presentation of high relevent words are 2.8 db. higher than thresholds obtained after neutral words. The corresponding figure for the day before a jump is 1.3 db. and for a day 2 weeks from a jump, .9 db. Control *S*s manifest a slight decreasing gradient as a function of the stimulus dimension, suggesting adaptation to the cues of parachuting. Tables 8, 9, and 10 indicate that all of the above effects are highly significant.

Trend analysis of the interaction between experimental condition (proximity to a jump) and stimulus dimension yields a significant difference (.001 level) in the linearity of the gradients as a function of time from a jump. Individual analyses carried out for each experimental condition reveal significant linear and cubic functions. The *F* ratios are more pronounced on the day of the jump than on the other 2 days of testing. In Figure 14 it can be seen that gradients become increasingly positively accelerated as the time of the jump approaches.

Again referring to Figure 14, it can be seen that there is a sharp increase in the threshold of parachutists for tones following neutral words from the control day to the day of a jump (1.82 db.). When tested by analysis of variance, the difference is significant at the .01 level. It may be concluded that parachutists on the day of a jump demonstrate auditory deficit independent of the effect of activation-arousing cues. On the day 2 weeks from a jump parachutists do not differ from control *S*s.

Turning to the effect of anxiety words on threshold for tones, it can be seen from Figure 14 that parachutists on each of the 3 days of testing have a considerably higher auditory threshold for tones following anxiety words than for tones following

TABLE 6

ANALYSIS OF VARIANCE OF BASAL CONDUCTANCE TAKEN AT THREE TIME INTERVALS DURING THE EXPERIMENT FOR CONTROL *S*S AT THREE TESTING SESSIONS

Source	*df*	*MS*	*F*
Between *S*s	26		
Sequence (Q)	2	3.82	.45
*S*s/Q	24	8.42[a]	4.73***
Experimental Condition (E)	2	3.26	.83
Q × E	4	1.70	.43
*S*s × E/Q	48	3.92[a]	2.20***
Time (T)	2	1.02	.28
Q × T	4	4.44	1.24
*S*s × T/Q	48	3.57[a]	2.00***
E × T	4	.79	.44
Q × E × T	8	3.68	2.06
*S*s × E × T/Q	96	1.78[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × E × T/Q interaction was used to evaluate all other error terms.

*** $p < .001$.

neutral words. Analysis of variance reveals that this effect is significant at the .001 level. While the greatest differential effect occurs on the day of a jump, a failure of the interaction of experimental condition (proximity to a jump) and stimulus dimension to reach significance, indicates that the greater steepness of gradient on the day of a jump than on the other days is not reliable. The trend nevertheless is in the direction of the hypothesis. Control *Ss* show the same effect as parachutists (significant at the .001 level) indicating, as for GSR, that anxiety cues, which are not specific to parachuting, produce similar effects in both groups.

Response Latency

Hypotheses involving response latency were generally confirmed. In Figure 15 it can be seen that at all three times of testing parachutists produce positive gradients as a function of the stimulus dimension, and that the gradient is higher and steeper on the day of the jump than on the day before a jump and least steep 2 weeks from a jump. The mean difference in reaction time for neutral and high words is 1.67 seconds on the day of a jump, .96 seconds on the day before a jump, and .63 seconds 2 weeks from a jump. No reliable difference in reaction time as a function of the stimulus dimension occurs for control *Ss*. Analyses of variance, shown in Tables 11, 12, and 13 indicate that all the above effects are significant.

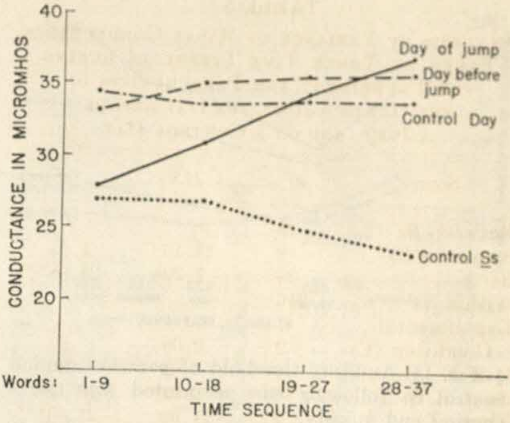


FIG. 13. Absolute level of conductance of parachutists and control *Ss* throughout the word list.

TABLE 7

ANALYSIS OF VARIANCE OF BASAL CONDUCTANCE TAKEN AT THREE INTERVALS DURING THE EXPERIMENT FOR PARACHUTISTS AND CONTROL *Ss*

Source	*df*	*MS*	*F*
Between *Ss*	53		
Groups (G)	1	39.40	5.06**
Ss/G	52	7.79[a]	7.02***
Within *Ss*	432		
Experimental Condition (D)	2	3.64	1.10
G × E	2	1.72	.52
Ss × E/G	104	3.31[a]	2.98***
Time (T)	2	6.51	3.15*
T × G	2	5.54	2.68
Ss × D/G	104	2.06[a]	1.85***
T × E	4	.89	.80
T × E × G	4	3.18	2.87*
Ss × E × T/G	208	1.11[a]	

[a] Error term for *MS*s above it up to the next error term. The *Ss* × E × T/G interaction was used to evaluate all other error terms.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

The trend analysis for the interaction between experimental condition and stimulus dimension indicates that the difference between gradients is associated with linear, quadratic, and cubic components. Separate analyses for each experimental condition reveal a significant linear, quadratic, and cubic function on the day of a jump; a significant linear and cubic function on the day before a jump; and a significant linear, quadratic, and cubic function on the day 2 weeks from a jump. All effects are more clearly pronounced on the day of a jump than on the day 2 weeks from a jump. While the pattern is less orderly than for GSR, Figure 15 reveals the same increasing positive acceleration of the curves as a function of decreasing time to a jump.

Turning to the results on reaction time to neutral words, Figure 15 clearly reveals that parachutists are more rapid responders than control *Ss*. Mean reaction time of parachutists to neutral words on a control day (1.16 seconds) is significantly lower (.01 level) than mean reaction time of control *Ss* (1.95 seconds). Increasing reaction time as a function of decreasing time to a

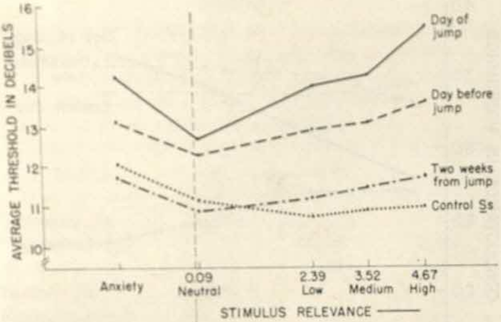


FIG. 14. Auditory threshold of parachutists and control *S*s following cues associated with parachuting and anxiety.

jump is in the predicted direction, but falls short of significance.

Figure 15 further shows that an increase in reaction time from neutral to anxiety words is demonstrated by all groups. However, there is a tendency for this effect to be greatest for parachutists on the day of a jump. Analyses of variance of these data indicate significance for the effect of dimension, but fail to indicate a differential effect of the dimension associated with groups or time of testing in relation to a jump.

Perception

Perception differs from sensation in that it involves the ascription of meaning to a stimulus. Using as a point of reference the objective meaning of the stimulus, as defined by consensus, perceptual responses can be scored as accurate or inaccurate. In the present study, a gross measure related to perceptual threshold was obtained by simply noting whether words were correctly or incorrectly perceived. Misperceptions could be unambiguously scored when *S* reported that he had not heard a word, or when later questioning of obviously inappropriate responses indicated that the stimulus word was mistaken for another, frequently with one that rhymed with it. Misperception in the word-association test could be scored in reference to approach and avoidance of cues associated with parachuting. In addition, a score of more general perceptual deficit could be obtained by attending only to neutral words.

Figure 16 presents number of misperceptions of words in a category, presented as a percentage of all words in a category. It is immediately apparent that parachutists on the day of a jump misperceive a high proportion of anxiety words and relatively few parachute-relevant words. Inspection of the frequency distribution indicates that there are a high number of cases with zero misperceptions accounting for markedly skewed distributions. Accordingly chi-square analyses of frequency data were resorted to.

Of 20 parachutists who, on the day of a jump, misperceive a different number of neutral words and words highly relevant to parachuting, 18 make fewer errors on the parachute words, which is significant at the .01 level. That this is not due to any intrinsic difference in the difficulty of the words is indicated in the finding that control *S*s do not differ in misperception of neutral and parachute-relevant stimuli. Thus, of the 23 control *S*s who produce an unequal number of misperceptions to the two categories, 12 produce more misperceptions to the neutral words and 11 to the words related to parachuting.

TABLE 8

ANALYSIS OF VARIANCE OF AUDITORY THRESHOLD TO A STIMULUS DIMENSION FOR PARACHUTISTS ON THE DAY OF A JUMP, THE DAY BEFORE A JUMP, AND ON A CONTROL DAY

Source	*df*	*MS*	*F*
Between *S*s	26		
Sequence (Q)	2	247.22	2.34
*S*s/Q	24	105.23[a]	159.44***
Within *S*s	297		
Experimental Condition (E)	2	213.79	15.74***
Q × E	4	8.99	.66
*S*s × E/Q	48	13.58[a]	20.57***
Dimension (D)	3	38.93	31.86***
Q × D	6	1.59	1.30
*S*s × D/Q	72	1.22[a]	2.03***
E × D	6	4.56	6.95***
Q × E × D	12	.55	.84
*S*s × E × D/Q	144	.66[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × E × D/Q interaction was used to evaluate all other error terms.

*** $p < .001$.

Consistent with findings on the day of a jump are findings on the other 2 days of testing: The ratio is 19 to 3 in favor of fewer misperceptions of parachute-relevant words than neutral words on the day before a jump ($p < .01$), and 19 to 4 on the day 2 weeks from a jump ($p < .01$). The hypothesis of differences between experimental condition and proximity to a jump is not upheld, although the tendency in Figure 16 is in the expected direction. It thus may be concluded that (*a*) parachutists perceive parachute-relevant words more readily than neutral words; (*b*) perception of parachute-relevant words by parachutists does not vary significantly as a function of experimental condition, although the tendency is in the predicted direction; and (*c*) control *S*s do not demonstrate differential perception of neutral words and words related to parachuting.

TABLE 9

ANALYSIS OF VARIANCE OF AUDITORY THRESHOLD TO A STIMULUS DIMENSION FOR CONTROL *S*S AT THREE TESTING SESSIONS

Source	*df*	*MS*	*F*
Between *S*s	26		
Sequence (Q)	2	222.90	1.82
*S*s/Q	24	122.54[a]	121.22***
Within *S*s	297		
Testing Sessions (T)	2	50.75	2.50
T × Q	4	18.47	.91
*S*s × T/Q	48	20.22[a]	2.19***
Dimension (D)	3	2.51	3.24*
D × Q	6	.99	1.27
*S*s × D/Q	72	.77[a]	.76
T × D	6	1.93	1.91
T × D × Q	12	.97	.97
*S*s × T × D/Q	144	1.01[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × T × D/Q interaction was used to evaluate all other error terms.

* $p < .05$.

*** $p < .001$.

TABLE 10

ANALYSIS OF VARIANCE OF AUDITORY THRESHOLD TO A STIMULUS DIMENSION FOR PARACHUTISTS AND CONTROL *S*S

Source	*df*	*MS*	*F*
Between *S*s	53		
Groups (G)	1	561.58	4.55*
*S*s/G	52	123.20[a]	148.43***
Experimental Condition (E)	2	28.80	1.72
G × E	2	235.73	14.15***
*S*s × E/G	104	16.66[a]	20.00***
Dimension (D)	3	15.53	15.22***
D × G	3	25.89	25.38***
*S*s × D/G	156	1.02[a]	1.22***
D × E	6	1.35	1.63
G × D × E	6	5.14	6.22***
*S*s × E × D/G	312	.83[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × E × D/G interaction was used to evaluate all other error terms.

* $p < .05$.

*** $p < .001$.

Turning to anxiety words, the indication in Figure 16 that parachutists misperceive more anxiety words on the day of a jump than any other day of testing and than control *S*s is in support of the hypothesis. The hypothesis that the same effect occurs on the other days, but to a lesser extent, is not supported. Of parachutists who produce an unequal number of misperceptions of neutral and anxiety words 2 weeks from a jump, 1 makes more errors on anxiety words, while 20 miss more neutral words (about 5%). On the day before a jump, corresponding figures are 1 and 8 (about 11%), and on the day of a jump, 13 and 12 (about 52%). It is apparent that there is generally a much greater likelihood for *S*s to miss neutral words than anxiety words in the present list, although the findings show a clear tendency of increased anxiety words missed with greater proximity to a jump. A comparison of number of parachutists who misperceive more anxiety words than neutral words on the day of a jump relative to the day before a jump yielded a difference significant at the .001 level ($\chi^2 = 12.04$), at the .01 level for day of jump versus day 2 weeks before a jump ($\chi^2 = 8.84$), and at the .05 level for day before a jump versus day 2 weeks from a jump ($\chi^2 = 6.50$). No significance was found for control *S*s or between parachutists and control *S*s in any single comparison. It may be concluded that parachutists on the day of a jump show a deficit for anxiety words, while on days more remote from a jump they are relatively sensitized to anxiety words.

With regard to misperception of neutral words, it was hypothesized that the number

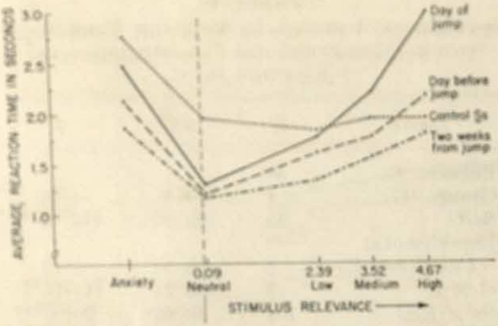


FIG. 15. Reaction time of parachutists and control *Ss* to a stimulus dimension and to anxiety words.

of misperceptions increases with closeness to a jump, as the result of a very general perceptual deficit. The results, when tested by chi square, fall short of significance. There is no significant difference between parachutists and control *Ss* on neutral words under any of the experimental conditions. In Figure 16 it can be seen that there is nevertheless a tendency for parachutists to misperceive more neutral words on the day of a jump than on the other occasions.

Figure 17 presents percentage of misperceptions for neutral words categorized according to the word they follow. It is evident that parachutists on the day of a jump produce a positive gradient of misperceptions as a function of the degree of relevance to parachuting of the preceding word. Of 22 parachutists who misperceive a different number of neutral words following words of high parachute relevance on the day of a jump and 2 weeks from a jump, all 22 produce more misperceptions on the day of a jump, which is significant at the .001 level. A similar analysis comparing the day of a jump and the day before a jump reveals that 18 out of 21 *Ss* misperceive more neutral words following relevant words on the day of a jump, which is significant at the .01 level. Of 13 *Ss* who misperceive an unequal number of neutral words following high-relevant words on the day before a jump and 2 weeks from a jump, 12 make more such misperceptions on the day before a jump, which is significant at the .01 level. Thus, when parachutists are used as their own controls, they demonstrate an increase in deficit associated with perception of neutral words following high-relevant words as time to an anticipated jump approaches. The same conclusion is supported when parachutists are evaluated relative to control *Ss*. There are 22 parachutists on the day of a jump who misperceive more neutral words following high relevant words, as compared to 3 control *Ss*. The difference is significant at the .001 level. When comparison is between parachutists on the day before a jump and control *Ss*, 12 parachutists and 1 control *S* show greater deficit to neutral words following high-relevant words. This difference is significant at the .01 level. Parachutists do not differ from control *Ss* 2 weeks from a jump. It may be concluded that a parachute-relevant word influences the perception of the immediately following non-parachute-relevant word to the extent of its own activation-producing effect on parachutists due to distance in time to a jump and to its relevance to parachuting. No effect is noted for control *Ss*.

A comparison between neutral words

TABLE 11

ANALYSIS OF VARIANCE OF REACTION TIME TO A STIMULUS DIMENSION FOR PARACHUTISTS ON THE DAY OF A JUMP, THE DAY BEFORE A JUMP, AND ON A CONTROL DAY

Source	*df*	*MS*	*F*
Between *Ss*	26		
Sequence (Q)	2	1.45	.92
Ss/Q	24	1.56[a]	.24
Within *Ss*	297		
Experimental Condition (E)	2	9.31	43.32***
Q × E	4	.33	1.52
Ss × E/Q	48	.22[a]	.03
Dimension (D)	3	16.70	98.71***
Q × D	6	8.52	.50
Ss × D/Q	72	.17[a]	.02
E × D	6	1.44	22.59***
Q × E × D	12	7.74	1.21
Ss × E × D/Q	144	6.38[a]	

[a] Error term for *MS*s above it up to the next error term. The *Ss* × E × D/Q interaction was used to evaluate all other error terms.

*** $p < .001$.

following high-relevant words with neutral words following neutral words reveals that the two differ. Of 19 parachutists who differed in number of misperceptions on the day of a jump and 2 weeks from a jump, all made relatively more misperceptions of neutral words following parachute-relevant words on the day of a jump, an outcome which is significant at the .001 level. For comparison of performance on the day before a jump and 2 weeks from a jump, there were 12 parachutists who produced a different number of misperceptions. All 12 of these *S*s made more misperceptions of neutral words following parachute-relevant words on the day before a jump ($p < .01$). When the comparison is between parachutists on the day of a jump and control *S*s, 19 out of 27 parachutists, and 2 out of 27 control *S*s misperceived more neutral words following words of high relevance than neutral words following neutral words, a difference which is significant at the .01 level. In the comparison between parachutists on the day before a jump and control *S*s, 12 out of 27 parachutists and 1 out of 27 control *S*s misperceived relatively more neutral words following high-relevant words than neutral words following neutral words, which is significant at the .01 level. It may be concluded that there is a selective increase in misperception of neutral words following words of high relevance for parachutists as time to an anticipated jump approaches.

TABLE 12

ANALYSIS OF VARIANCE OF REACTION TIME TO A STIMULUS DIMENSION FOR CONTROL *S*s AT THREE TESTING SESSIONS

Source	*df*	*MS*	*F*
Between *S*s	26		
Sequence (Q)	2	10.84	4.71*
*S*s/Q	24	2.30^{a}	13.53***
Within *S*s	297		
Testing Sessions (T)	2	.54	.85
T × Q	4	.67	1.05
*S*s × T/Q	48	.63^{a}	3.70***
Dimension (D)	3	.26	1.95
D × Q	6	.13	.96
*S*s × D/Q	72	.13^{a}	.76
T × D	6	.14	.84
T × D × Q	12	.23	1.42
*S*s × T × D/Q	144	.17^{a}	

[a] Error term for *MS*s above it up to the next error term. The *S*s × T × D/Q interaction was used to evaluate all other error terms.
* $p < .05$.
*** $p < .001$.

TABLE 13

ANALYSIS OF VARIANCE OF REACTION TIME TO A STIMULUS DIMENSION FOR PARACHUTISTS AND CONTROL *S*s

Source	*df*	*MS*	*F*
Between *S*s	53		
Groups (G)	1	5.12	2.26
*S*s/G	52	2.25^{a}	18.75***
Within *S*s	594		
Experimental Condition (E)	2	7.04	16.36***
G × E	2	2.82	6.57**
*S*s × E/G	104	.43^{a}	3.57***
Dimension (D)	3	8.58	57.61***
D × G	3	8.38	56.21***
*S*s × D/G	156	.15^{a}	1.20***
G × D × E	6	.62	5.23***
*S*s × E × D/G	312	.12^{a}	

[a] Error term for *MS*s above it up to the next error term. The *S*s × E × D/G interaction was used to evaluate all other error terms.
** $p < .01$.
*** $p < .001$.

Memory

Each list of 37 words contains 3 words in five categories that are scored on all measures: 5 practice words and 17 neutral buffers. In the memory task *S*s were asked to recall all word stimuli together with their respective responses. At times *S*s were able to remember only the stimulus word or the response word. On other occasions, a wrong response was produced to the right stimulus and vice versa. Inspection of these different kinds of errors failed to reveal functional differences. Accordingly the data were pooled into three indexes of recall: correct recall of neutral stimuli and responses, correct recall of parachute-relevant stimuli and responses, and correct recall of anxiety stimuli and responses. A division of responses according to the degree of parachute relevance of the stimulus word was found to add nothing and was discarded.

Figure 18 presents the findings on memory. It may be seen that while there is a

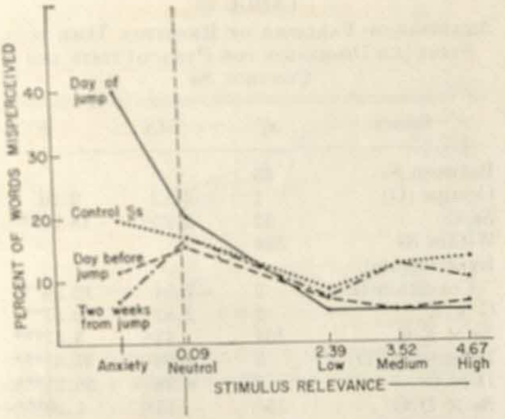


FIG. 16. Percentage of misperceptions by parachutists and control *S*s of words on a stimulus dimension and anxiety words.

general tendency to recall more parachute-relevant than neutral words, this tendency is more pronounced for parachutists than for control *S*s. Analyses of variance indicate the difference between parachutists and control *S*s in recall of parachute-relevant words relative to neutral words is significant at the .05 level. Figure 18 also shows a tendency for parachutists to recall more parachute-relevant words on the day of a jump than on the day before a jump, or on the day 2 weeks from a jump, in that order, as had been hypothesized; but this difference was not significant. It may further be seen in Figure 18 that the trend of differential recall of neutral and anxiety words is opposite from what had been hypothesized. There is, however, no significant effect due to experimental condition or groups, all groups demonstrating a constant bias in favor of recall of anxiety words. It may be concluded that parachutists and nonparachutists recall more parachute-relevant and anxiety words than neutral words, and the effect of closeness to a jump increases the tendency for recall of parachute-relevant words for parachutists. While this finding supports the hypothesis for sensitization toward parachute-relevant material, the hypothesis for defense against anxiety-relevant material was not confirmed.

To evaluate generalized memory deficit as a measure of performance, it was decided to use a score representing the recall of the 22 "neutral" buffer words rather than only the 3 neutral words. The hypothesis that parachutists' level of recall would decrease as the time of a jump approached was substantiated. The average number of neutral stimuli and responses correctly recalled was 3.88 on the day of a jump, 5.66 on the day before a jump, and 8.18 2 weeks from a jump. Analysis of variance reveals the difference to be significant at the .01 level. Control *S*s recall on an average 7.24 neutral stimuli and responses per session and do not differ significantly over sessions. Comparing parachutists to control *S*s, the difference between groups and experimental condition is significant at the .01 level. It may be concluded that, so far as neutral words are concerned, parachutists in the control condition do not differ from control *S*s, but show increasing memory deficit as the time of a jump approaches.

Association

Although there was some indication of response deficit to correctly perceived stimuli, such as the failure to produce a response, the repetition of a stimulus word, the repetition of a response word, or the production of a neologism, such responses were relatively rare. Of these, the most common was the repetition of a response

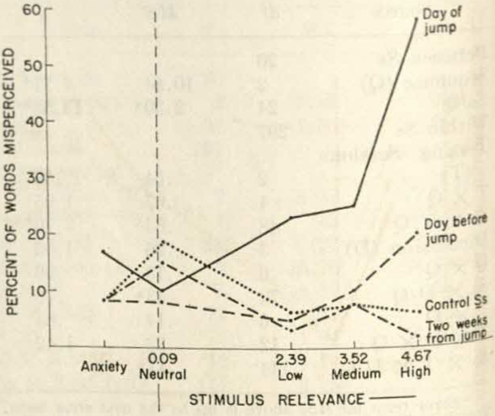


FIG. 17. Percentage of misperceptions of parachutists and control *S*s of neutral words following words on a stimulus dimension and anxiety words.

word, as illustrated in the following example of a record of a parachutist on the day of a jump:

SKYDIVER - JUMP; HUNGRY - JUMP; RADIATOR - HOT; HAPPY-PARACHUTE; BICYCLE-JUMP, damn it, I can't get away from it!

Unfortunately there were not enough such low-level reponses to allow for statistical analysis, although the trend was clearly in the direction of parachutists producing more such responses on the day of a jump than at any other time.

The number of parachute-relevant responses produced was investigated as a measure of approach. Since there was no evidence that stimulus words at different points along the stimulus dimension were functionally dissimilar, responses to all stimuli were combined into an overall score. Parachutists on the day of a jump produce an average of 12.8 parachute-relevant associations, parachutists on the day before a jump 7.9 such responses, and parachutists 2 weeks from a jump 6.1 such responses. Control *S*s produce the least number of parachute-relevant responses, with a mean of 4.8. Differences between parachutists at the three testing sessions, and between parachutists and controls, were tested by analysis of variance and were found to be all highly significant. It may be concluded that parachutists, irrespective of testing session, produce more parachute-relevant associations than control *S*s and that the number of parachute-relevant associations is directly influenced by proximity to a jump.

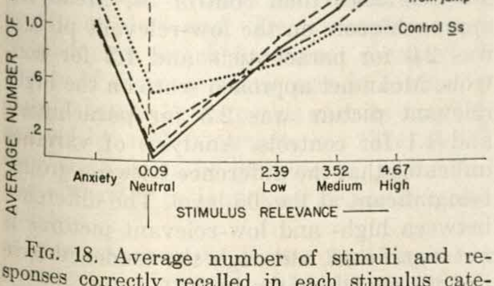


Fig. 18. Average number of stimuli and responses correctly recalled in each stimulus category by parachutists and control *S*s.

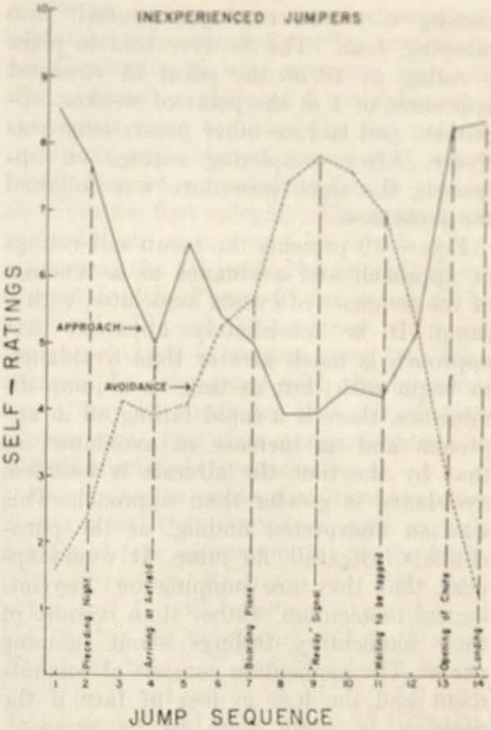


Fig. 19. Self-ratings of approach and avoidance as a function of the sequence of events leading to and following a jump.

Self-Ratings of Approach and Avoidance

Shortly after landing from a parachute jump, parachutists were asked to rate their feelings of approach and avoidance at each of the following points along a sequence of events associated with parachuting: (*a*) last week, (*b*) last night, (*c*) this morning, (*d*) upon reaching the airfield, (*e*) during training period before jump, (*f*) getting strapped, (*g*) boarding aircraft, (*h*) during ascent, (*i*) at the "ready" signal, (*j*) stepping outside, (*k*) waiting to be tapped, (*l*) freefalling, (*m*) after chute opened, (*n*) immediately after landing. *Approach* was defined as "looking forward to the jump, wanting to go ahead, being thrilled by the prospect of jumping"; *avoidance* as "wanting to turn back and call the jump off, questioning why you ever got yourself into jumping, fear." The *Ss* were told to place a rating of 10 at the point of strongest approach, of 1 at the point of weakest approach, and to rate other points comparatively. After completing ratings on approach, the same procedure was followed for avoidance.

Figure 19 presents the mean self-ratings of approach and avoidance as a function of the sequence of events associated with a jump. It is immediately apparent that approach is much greater than avoidance, to begin with; but as time to a jump diminishes, there is a rapid falling off in approach and an increase in avoidance, so that by the time the aircraft is boarded, avoidance is greater than approach. This was an unexpected finding, as the parachutists, after all, do jump. It would appear that they are jumping on "psychological momentum" rather than because of their momentary feelings about jumping per se. The momentum consists of commitment and the fear of loss of face if the decision to jump were to be reversed. Avoidance continues to become increasingly greater than approach, reaching a peak when the signal is given to get ready to jump. It is noteworthy that the greatest avoidance reaction is not associated with the moment of jumping, but with a point that involves a final commitment to jump. When the parachutist is standing out on the step above the wheel, waiting to be tapped, where jumping is a foregone conclusion, there is a decrease in avoidance. At Point 12 in the sequence, where *S* is freefalling and objective danger is maximal, approach again exceeds avoidance.

Content of Thematic Responses

It will be recalled that the thematic apperception test was administered only on the day of a jump, and permitted therefore no comparison on the time effect. Criteria for scoring were the same as the ones used by Fenz and Epstein (1962) and are described in detail in that study. Each story was typed on a separate card; and cards were coded and shuffled; and stories scored blindly. All stories to the same picture were then sorted into separate columns representing different weights, allowing for ready comparison within and between stories in each column.

Approach Content. A score of the net approach attempted to assess the overall approach or avoidance to parachuting manifested in a story. Judges were told to make a global estimate of the degree to which the storyteller's thoughts centered on parachute-relevant activities and were favorable or unfavorable to jumping and to then assign a weight of 0 to stories suggesting no approach to parachuting, a weight of 4 to stories most strongly indicating approach to parachuting, and to give relative intermediate weights. They were told to assign a positive reference to parachute-relevant activities, such as flying ("I wish I could be up there with the pilot"), a higher score than a simple reference to a boy admiring the flight of birds ("this boy is gazing in wonderment at the birds, how they go so freely with the wind"), while a positive reference to parachuting itself ("he is looking forward to the excitement and adventure of the free fall") should receive a yet higher score.

With the above consideration in mind, the stories were independently scored by two examiners. For the low-relevant picture an interscorer reliability coefficient of .83 and for the high-relevant picture a coefficient of .83 were obtained. Where disagreement between scores was within two points, ratings were averaged; where greater than two points, differences were resolved by discussion.

As had been hypothesized, parachutists showed greater approach to parachuting in their stories to pictures of both high and low relevance than control *S*s. Mean net approach score on the low-relevant picture was 2.0 for parachutists and 1.2 for controls. Mean net approach score on the high-relevant picture was 2.3 for parachutists and 1.1 for controls. Analysis of variance indicates that the difference between groups is significant at the .05 level. The difference between high- and low-relevant pictures is not significant, although the tendency is in the direction hypothesized.

Fear of Parachuting. It was noted that

Figure 9e, which is highly relevant to parachuting, elicited many fear responses, such as, "He says to himself, what the hell am I doing up here anyway?" "He looks worried . . . What will happen . . . and thinks: Am I gonna die?" "The guy is scared stiff. He is biting his lips, wondering if he will have the courage. . . ." It was decided to score these responses on a continuum of fear of parachuting, and it was found that responses could be discriminated along five points. The method of scoring followed the same procedure as for approach content. Judges were told to rate responses on a 5-point scale, with 0 indicating an absence of fear ("he is not scared at all") and 4 ("I am not going out there") a high degree of fear. An interscorer reliability coefficient of .92 was obtained. Final scores consisted of the average ratings of both examiners.

Expression of fear of parachuting yielded a mean score of 1.35 for parachutists, and 1.92 for controls. The difference tested by analysis of variance is significant at the .05 level. It may be concluded that parachutists express less fear than control *S*s to a picture of high relevance to parachuting.

Defense Mechanisms. If it could be shown that parachutists express less fear to parachute-relevant stimuli, and more fear to stimuli unrelated to parachuting than control *S*s, a strong point for the use of a defense mechanism of displacement by parachutists could be made, and for their need to control and exhibit fear.

Figures 9b and 9f elicited a considerable number of references to fear unrelated to parachuting. To Figure 9b, which shows two boys running, some examples of fear responses were: "He is scared stiff of being beaten up . . . " and "He is running away from danger which he cannot escape. . . ." Figure 9f shows a young man in front of a stately home. Fear to this picture was expressed by such comments as: "The closed door is an omen of disaster . . ." and "The boy is scared; he is picking up his girl, and wonders. . . ."

Fear responses to Figure 9b were rated by two examiners on a 5-point scale and yielded an interscorer reliability coefficient of .79. An average fear score for parachutists of 1.6 was obtained and for control *S*s, 1.0. Combining *S*s who expressed fear, and who obtained a score of 1 or higher, and comparing them to those who did not express fear, and obtained a score of 0, it was found that 14 parachutists and 8 control *S*s fell in the first category. The comparison yielded a chi square of 4.50, which is significant at the .05 level. Not enough *S*s expressed fear to Figure 9f for scores to be analyzed, although the tendency again was for parachutists to express more fear than controls. In accordance with the hypothesis, it may be concluded that parachutists express more fear than nonparachutists to stimuli unrelated to parachuting.

Examining the unidentified pool of stories it was noted that there were nine explicit denials of fear of parachuting to the high-relevant picture. Examples are: "He does not worry. It will be a real experience in his life, but his sweetheart and mother are afraid for him . . . " and "He looks so relaxed, he is not scared at all. . . ." All such stories were found to be produced by parachutists, which is significant beyond the .01 level.

In summary, parachutists on the day of a jump produce strong approach responses to parachuting in their Thematic Apperception Test stories, explicitly deny fear of parachuting, and produce an inordinate number of fear responses unrelated to parachuting.

DISCUSSION

The mental state of a parachutist making his first or second jump is very similar to that of a psychiatric patient who is seriously troubled by some very real problem. While an experienced jumper is, in most cases, capable of exerting control over his emotions and actions, a beginning jumper, at the time of the jump, is overwhelmed by his fears. Yet, at this point, to back out and give up is unacceptable. The result is an acute conflict resulting in a deterioration of performance. Much of the training in sport parachuting is directed toward making the *S* react by reflex, since experience has shown that the beginning jumper is apt not to follow the simplest instructions, such as, waiting for the jumpmaster's signal to jump, keeping a count of time elapsed until the chute opens, and assuming a proper position during the period of free fall. The same parachutists had no difficulty in performing the operations on the ground during training exercises. Thus it is apparent that the deficit exhibited in the air can be attributed to stress. To keep the deficit from producing disaster, the first few jumps are "static line" jumps in which the ripcord is automatically pulled for the jumper by the jumpmaster who remains in the aircraft. A report of the Parachute Club of America (Moore, 1963) indicates that chances of malfunctions are 10:1 in favor of the first free fall (when the jumper for the first time pulls his own ripcord) which usually comes between the fifth and eighth jumps, and that failures are essentially due to a lack of cognitive control on the part of the jumper. The condition induced by a forthcoming jump for a beginning parachutist may be viewed as a miniature abnormal state. It is in this light that the findings become meaningful in their implications for the understanding of maladaptive behavior in the behavior disorders. The high reliability that was found for some of the measures indicates that they are completely adequate for measuring individual differences, and that they thus have important implications for the development of diagnostic methods for determining areas of conflict and stress. The fact that the findings on various measures could be meaningfully integrated about a few key concepts also points to the fruitfulness of the theoretical approach. Central to the theoretical model presented in this study is the assumption that conflict can be measured by its correlates of physiological activation, performance deficit, and directionality of verbal content. The remainder of this paper is devoted to a discussion of the relationship of each of these response systems to stress and conflict.

Activation

Physiological activation has been reported in the literature to be indicative of changes in mental states. In a study by Landis and Hunt (1935), *Ss* were asked to describe subjective feelings during a period of stimulation. When the GSRs were sorted according to the mental states reported, it was found that tension, fear, and confusion, in that order, produced the strongest GSRs. These three feelings are highly descriptive of the mental state of a novice parachutist immediately prior to jumping. The hypotheses on activation in the present study assumed that such feelings as tension, subsumed under the concept of activation, would increase as a function of a decrease in time to the approaching event and of an increase in cues associated with the event. The hypotheses were clearly supported using GSR as the measure of activation.

The reliability and validity of GSR as a measure of activation, assuming the covariance of the latter with time to a jump and relevance of cues to parachuting, were clearly demonstrated. It will be remembered that so far as the cue dimension is concerned, Luria (1932) used a somewhat similar technique. He characterized stimuli in a word-association test as indifferent, doubtful, and critical as to their relationship to a specific situational conflict and observed that behavioral disorganization varied in a direct manner, according to the degree of relationship of the stimulus to the area of conflict. The use of projective tests containing dimensions associated with an area of conflict was also reported by Epstein and Fenz (1962). Whereas Luria, however, had used motor activity as the measure which most clearly corresponds to the concept of activation, as used in this paper, Epstein and Fenz used GSR. Luria's measure of motor behavior is actually a combined index both of activation and performance deficit, each of which is measured separately in the study by Epstein and Fenz and in the current study. The present findings on GSR clearly corroborate the earlier findings. As in the previous work, parachutists produced gradients of GSR to cues of increasing relevance to parachuting under all testing conditions, and for each individual the gradient was steeper on the day of a jump than on a day 2 weeks from a jump. In like manner the average gradient of parachutists on the day of a jump was steeper than on the day before a jump, but in this case the results were not consistent for all *Ss*. Obviously there are factors other than proximity to a jump which influence steepness of gradients. This is well illustrated by one *S* who produced a steeper gradient on the day preceding a jump than on the day of a jump. (His order of testing was such that the first test was 2 weeks before a jump, the second test on the day of a jump, and the third the day before a jump.) Questioning revealed that after the second testing and just prior to the jump itself, he witnessed several minor malfunctions by other parachutists and expressed apprehension and doubt about his ability to cope with similar situations. Nevertheless, he made the jump. On the third testing session, he produced a very steep gradient. While he said he would make the next jump, which was scheduled for the following day, he never did; nor was he seen at the airfield thereafter. Similar incidents provide convincing evidence that GSR is a sensitive measure in differentiating intensity of reaction to stress both between and within individuals.

A measure of activation which has received less attention in the literature than GSR is the absolute level of conductance. There is reason to believe that absolute conductance varies with general state of alertness or activation (Woodworth & Schlosberg, 1954). Several studies have reported that activation is high when *S* is alert and low when he is relaxed (Farmer & Chambers, 1925; Freeman & Darrow, 1935; Richter, 1926). Darrow and Heath (1932) investigated the changes in absolute level of conductance that occurred during the course of an experiment. They found that conductance was high when *S* was being harnessed in the recording apparatus and rose yet higher as he awaited the onset of the experiment. Davis (1934) reported a rise in absolute conductance throughout the course of an experiment which involved rapid adding under conditions of distraction by a loud noise. Duffy and Lacey (1946) reported a fall in absolute conductance in an experiment involving the determination of *S*'s lower threshold for sound. The findings of Davis and Duffy and Lacey might seem to be contradictory, if one were not to take into consideration that whether absolute conductance rises or falls depends on how *S* perceives and reacts to the task at hand. This was clearly demonstrated in the present experiment where, after an initial rise in conductance while waiting for the task to start, control *Ss* manifested a drop in conductance, while parachutists on the day of a jump demonstrated a sharp rise. The rise was less marked on the day before a jump, and on the day 2 weeks from a jump there was even a slight drop in conductance from the beginning to the end of the experiment. It follows that the rise in absolute conductance during the testing session provides an index of degree of disturbance of the jumper in reacting to cues associated with parachuting. The implication for locating areas of psychopathology are that in a parallel clinical situation, an area of conflict, if tapped by appropriate stimuli, will overrule the effects of adaptation and reveal itself by a rise in conductance during testing. It should be noted that out of the 27 parachutists tested on the day of a jump, 23 manifested a rise in conductance throughout the session, indicating that the reliability should be sufficient for measuring individual differences in a clinical situation. It will be remembered that the initial level of conductance was lowest on the day of a jump and highest on the day remote from a jump. As a result of stimulation, parachutists showed a rise in conductance, the sharpest rise occuring on the day of a jump. The findings on initial level of conductance can be interpreted as revealing a general state of inhibition produced by mounting stress. Since they were contrary to the original hypothesis, they will need to be replicated. If verified, they point to the importance of considering different types of inhibition, at least to a distinction between a general inhibition which produces an overall blanketing effect, and inhibition specific to particular stimuli.

Performance Deficit

It was assumed that high levels of activation would produce performance deficit, and the findings were consistent with this assumption. Whether the autonomic nervous system causes changes in the central nervous system, as Luria (1932) has suggested, or whether the two systems are independently influenced by a third source, or both, a significant correlation between performance deficit and autonomic reactivity under conflict is of considerable significance in its own right.

There is reason to believe that the relationship of level of performance to activation can best be described by an inverted U-shaped curve (Malmo, 1959). That is, up to a point, which differs according to the nature of the performance measure, level of performance increases with increasing activation, but beyond that point performance level declines. Thus the qualification was made in the present study that an increase in activation, whether produced by cues or time, would produce deficit, provided that total activation was of sufficient magnitude; and it was assumed that such was the case for a novice parachutist.

A number of studies have related autonomic reactivity to performance. Thorne (1934) and Hartmann (1934) found an inverse relationship between autonomic reactivity and sensory acuity. Trehub (1954) reported a relationship between sensory deterioration and states of tension. He obtained continuous recordings of GSR and visual thresholds during a session in which conflict-arousing verbal stimuli were presented. A direct relationship was found between magnitude of GSR and sensory threshold. Silverman, Cohen, and Shamavonian (1959) investigated the relationship between GSR and performance in *S*s exposed to different degrees of stress in a human centrifuge. They found that when an increase in nonspecific GSRs suggested moderate alerting, psychomotor performance improved; but when hyperalerting was indicated by a further increase in nonspecific GSRs, performance declined. In other experiments reported in the same paper, the authors demonstrated variations in perceptual threshold as a function of "level of arousal" as indicated by nonspecific GSRs. Eysenck (1960) compared a neurotic and a normal group on a test of auditory flicker fusion. He found that the neurotics had lower flicker-fusion rates than normals. In other work he demonstrated that neurotics produced greater GSRs than normals. Taken together, the findings suggest a correlation between rate of flicker fusion and GSR productivity.

In the present study the correlation coefficients between magnitude of GSR and auditory threshold in individual records ranged from .37 to .58, with a mean of .46 for 10 randomly selected records of parachutists tested on the day of a jump. On the day of a jump, 14 *S*s produced a gradient for sensory acuity, and 27 *S*s produced a gradient for GSR. Thus, both measures covary with activation, whether the latter is manipulated by cues or time. However, while a positive relationship between the two measures is indicated, auditory threshold would appear to be a less reliable measure of conflict than GSR. It is likely that the specific effect of stimulus cues on auditory threshold was to a great extent overshadowed by the effect of temporal proximity, since the parachutists exhibited a marked overall rise in threshold on the day of a jump.

A relationship between autonomic reactivity and reaction time has been widely discussed in the literature. Lanier (1941) reported that affect-laden words influenced length of reaction time in the same direction and to about the same extent as GSR. Hathaway (1929) reported a correlation of .60 between GSR and reaction time, while Hunt and Landis (1935) found no relationship. Peterson and Jung (1907) reported that only in certain cases is there a clear relationship. Findings by Epstein and Fenz (1962) indicated a marked relationship between the two measures. It is thus apparent that the relationship between GSR and reaction time is not a direct and simple one. Assuming that the curve for performance is inverted U shaped (Malmo, 1959), the relationship between reaction time and activation would depend, in part, on the nature of the task and, in part, upon the degree and range of activation sampled. In the present study, the correlation coefficients between GSR and reaction time in individual records ranged from .56 to .83, with a mean of .71 for 10 randomly selected records of parachutists tested on the day of a jump. It is especially noteworthy that 27 out of 27 *Ss* produced gradients both for GSR and reaction time on the day of a jump. The high reliability of the findings with reaction time as a measure of performance deficit under states of high emotionality and conflict indicates its adequacy when used in conjunction with a scaled stimulus dimension for measuring both individual differences and changes within the same person.

Reaction time might well be considered to be of special significance to a parachutist: leaving the aircraft at the right moment, assuming the proper free-fall position, detecting malfunctions, and determining whether or not to pull the ripcord of the reserve chute, are all tasks in which time is a key element. From the altitude at which a beginning jumper leaves the aircraft, in the case of a total malfunction, he has only 16 seconds before hitting the ground. During these 16 seconds only the first 10 seconds are considered safe for activating the reserve chute, to permit its full deployment and to slow down the jumper's rate of descent enough to prevent serious injury. It might thus be suspected that reaction time, being of specific significance to a parachutist, should produce less deficit than other measures of performance. The evidence, nevertheless, is contrary and suggests that the novice jumper is not able to cope adequately even with performance which may be directly related to his own safety.

Findings similar to those for reaction time were obtained for measures of perception, memory, and association. These findings showed that there is a clear relationship between GSR and perceptual deficit. It will be remembered that the present study measured perceptual deficit with respect to the three scaled neutral words following words relevant to parachuting, and that results substantiated hypotheses. In Jung's (1919) experiments on word association a similar phenomenon was stated. Erroneous reproductions, especially when given after a prolonged reaction time, and a general failure of reproduction were used as indicators of "complexes." Jung notes that such failures often occurred not at the complex connected stimulus word (critical word), but at the word following it (postcritical stimulus). Similar findings were reported by Luria (1932). He found that postcritical reactions (which are all to neutral stimuli) exhibited delayed reaction time and other signs of disturbance. What is probably the case is that the rise in activation produced by the stimuli related to an area of conflict extends beyond the interstimulus interval. Moreover, the response to the critical stimulus may start the *S* ruminating and wondering about its adequacy and what it may reveal. Thus, he is caught quite unprepared at the presentation of the next stimulus. A further possibility is that the new stimulus serves as a disinhibitor of the excess tension which was activated but not released in the response to the critical word.

Findings on memory for neutral stimuli and responses and on deficit in associative responses augment and corroborate the general findings on deficit in performance under states of tension. Some parachutists, when asked on the day of a jump to recall words on the association test immediately after having taken it, were unable to recall a single word. An average recall of less than 4 out of 22 neutral and buffer words on the day of a jump versus more than 8 words on the day 2 weeks from a jump is a clear indication of deficit under conditions of stress.

The findings on the influence of stress and conflict on performance find wide corroboration in the clinic. Serious personal conflicts are known to produce generalized inadequate and inefficient behavior, since much of the patient's time and effort are consumed in coping with the area of anxiety. The present study suggests that generalized performance deficit is a promising measure for detecting areas of conflict.

Directionality of Response

It may be assumed that self-ratings of approach and avoidance, to a large extent, are simply different ways of rating the same thing, namely, net approach or net avoidance. Thus, the gradients of self-ratings cannot be taken as independent representations of the underlying approach and avoidance response tendencies, both of which are assumed to have positive slopes. It will be recalled that while self-ratings of avoidance had a positive slope, self-ratings of approach had a negative slope, varying in reciprocal relationship to the self-ratings of avoidance. It is nevertheless revealing that at a time remote from a jump self-ratings of approach were greater than avoidance, while close to a jump the position of the two ratings reversed. From this it follows that the inferred independent avoidance gradient rises more steeply than the inferred independent approach gradient with approaching time to a jump. These ratings clearly conform to the observation of beginning jumpers who generally appear much more reluctant to jump when out on the airfield than when discussing their forthcoming jump with friends back home. The findings also indicate that *Ss* tested in the present study were aware of this fact and were able to give a candid evaluation of their feelings. The consistency found in these reports from one *S* to another points to the reliability of the phenomenon. Of additional interest, the point of maximum avoidance was found to occur at the signal to get ready to jump rather than during freefall, the point of maximum danger, attesting to the role of commitment in influencing anxiety.

How do parachutists deal with their feelings of approach and avoidance in projective tests? In the present study it was found that novice parachutists expressed on the Thematic Apperception Test very much the opposite of what they said they felt on self-ratings, particularly in regard to fear. When asked to evaluate their feelings, they admitted fear about parachuting on the day of the jump, while, on the other hand, on the Thematic Apperception Test, they strongly denied fear about parachuting. In an analogous manner they reported a diminished desire to jump as the time to a jump approached and demonstrated an increase in approach responses on the Thematic Apperception Test. The same reaction tendencies as exhibited on the Thematic Apperception Test, i.e., approach to parachute-relevant cues and avoidance of anxiety cues, were exhibited in a variety of response measures on the word-association test. Approach to parachute-relevant cues was revealed in a decrease in perceptual threshold for parachute-relevant words, in an increase in selective recall of parachute-relevant stimuli, and in an increase in associations with parachute-relevant content, with greater proximity to a jump. It is as if the parachutist, despite a general state of high tension and an additional increase in tension brought about by thoughts of parachuting, nevertheless forces himself to concentrate on cues related to parachuting. Avoidance of fear was exhibited in the large number of anxiety words misperceived on the day of a jump, in addition to the denial of fear of parachuting in the Thematic Apperception Test. It is as if the parachutist closed his ears to what he did not want to hear and denied the fear that he wished were not there. The projective tests gave him the license to order the world in terms of his wishes, while the self-ratings demanded a realistic appraisal. The novice jumper avoids thinking, i.e., attempts to inhibit his fear about parachuting and forces himself to think about the thrills and excitement and the prestige of making a jump, because of the adaptive significance of such reactions once the decision to jump has been made.

What are the theoretical implications of these findings? When Miller's (1944) models of conflict and displacement (Miller, 1951) were first applied to projective techniques, it was hypothesized (Epstein & Smith, 1956; Miller, 1959) that conflict is indicated by an increase in conflict-relevant responses to cues of low relevance and a decrease of conflict-relevant responses to cues of high relevance. Recent studies by Epstein and Fenz (1962) and Fenz and Epstein (1962) yielded results which were opposite to this hypothesis, namely, approach responses to stimuli of high relevance increased more than to stimuli of low relevance as time to the jump approached. Trying to find an explanation for these findings, it was reasoned that projective tests measure verbal responses that reflect expressive and inhibitory tendencies, rather than approach and avoidance per se, and that the gradient of inhibition is steeper than the gradient of expression. In the present experiment these hypotheses derived from previous work on approach and avoidance as applied to projective tests were tested. As shown above, the hypotheses were generally substantiated. Thus, in the projective tests, the conflict that was measured was not one of approach and avoidance but of expression and inhibition of fear, with the inhibition exhibited in responses to cues of high relevance and the expression in responses to cues of low relevance. Findings on inhibition and denial of fear to cues of high relevance, which gains expression in reaction to cues of low relevance, obviously do not reflect the parachutist's conflict about whether to jump or not to jump, but indicate instead how he copes with the fear.

While the above comments point to the adaptiveness of sensitization and defense in a particular stressful situation, it is a well-known clinical fact that expressive sensitization and defense are hallmarks of maladaptive behavior. Thus, deficit is produced not only by excessive general activation, but also by excessive focus on one source of cues to the exclusion of other sources of information. Clinical data on fixation, compulsions, and paranoid delusions illustrate how a thought may become so central and persisting in a person's thinking that it obfuscates and pervades all others. A pertinent case, which illustrates excessive sensitization to the point of producing marked performance deficit, was included in the present sample. On the day remote from a jump this parachutist's responses on the word-association test were totally adequate. On the day of the actual jump, however, parachuting had become so central to his thinking, that he could not keep from giving parachute-relevant responses even when totally irrelevant to the eliciting stimulus, such as SKYDIVER to CHAIR and JUMPMASTER to STOVE. This provides a good example of performance deficit due to excessive focus of interest, as opposed to the generalized deficit already discussed. During ascent in the plane this same *S* fell asleep and had to be aroused forcefully to get ready for the jump. Since the jumper was sitting at the open door of the plane, the blast of the wind and the noise of the engine were hardly conducive to sleep. The behavior of this jumper under stressful conditions may be interpreted as maladaptive defensiveness. It suggests that a failure of inhibition and control at high levels of stress or activation, as illustrated in the inappropriate parachuting intrusions on the word-association test, is apt to be followed by excessive and overly generalized inhibition at yet higher levels of stress. Perhaps such a person is not able adequately to modulate his inhibition of fear and thus exhibits an all-or-none reaction.

Broverman and Lazarus (1958) speak of an opposite form of loss of cognitive control, which is associated with an inability to concentrate on the important aspects of a task rather than becoming overly focused. He developed a method for measuring cognitive control in term of the individual's ability to resist the influence of interfering or irrelevant stimuli. Findings in the present study indicate that under moderate states of tension parachutists generally showed adaptive focusing responses as indicated by sensitization to cues associated with parachuting and defensiveness to cues associated with fear of parachuting. Under yet higher states of tension, as when getting strapped into the harness, during take-off, and while gaining altitude, jumpers often asked strangely irrelevant questions. An experienced jumper has learned to interpret these as requests for reassurance and to answer them most effectively with a reassuring pat on the back. In accordance with Broverman and Lazarus' (1958) findings, it appears that under this high state of tension an *S* is no longer able to resist the influence of irrelevant stimuli, thereby showing indexes of lack of control.

It may thus be concluded that stress influences performance via several different mechanisms. Being sensitized to a task may facilitate performance in that task to the detriment of performance in other tasks. If the focus is to be restricted to a particular aspect of a task, other aspects of the task will suffer, and performance will decline. Thus, the increase in focus activity at high levels of stress can reach a point of deficit. An opposite phenomenon, that of inadequate focus and an attendant inability to respond differentially to relevant and irrelevant cues is also apt to occur at the highest levels of stress.

SUMMARY

The present study investigated the relationship between stress experienced in a real-life approach-avoidance conflict situation, and physiological activation and performance. A three-dimensional model permitted evaluation of conflict in relation to both temporal and cue dimensions.

Twenty-seven sport parachutists were tested three different times in relation to a jump. The test consisted of stimuli scaled along a dimension of increasing relevance to parachuting. GSR and absolute level of conductance measured physiological activation, while performance was sampled over a wide range of responses, varying from simple sensory functions, at one extreme, to complex cognitive processes at the other.

A direct relationship between physiological activation and performance deficit was noted. As had been hypothesized, parachutists produced increasingly steep gradients of GSR to parachute-relevant and anxiety words with increasing proximity to a jump. Parachutists also produced gradients for auditory threshold of pure tones following words on the dimension and for reaction time to these words. Parachutists further became generally more reactive on the day of a jump as compared to days more remote from a jump and showed a corresponding deficit in general performance.

Absolute level of conductance was measured prior to, and during, testing. Prior to testing it was lowest on the day of a jump and highest on a day remote from a jump, but during testing itself the relative position reversed itself. This finding was explained as reflecting a general state of inhibition under states of tension which nevertheless is not capable of warding off the effect of intense stimulation.

Self-ratings of approach and avoidance indicated that novice parachutists are afraid of jumping, and that fear becomes greater as the time of the jump approaches. On the Thematic Apperception Test, parachutists nevertheless strongly denied their fear of jumping. This was viewed as an adaptive way of handling a fear-provoking situation which parachutists feel obliged to face. In addition, parachutists showed sensitization to parachute-relevant cues, as reflected in measures of perception, memory and association, and defense against anxiety-provoking cues.

REFERENCES

BROVERMAN, D. M., & LAZARUS, R. S. Individual differences in task performance under conditions of cognitive interference. *J. Pers.*, 1958, **26**, 94–105.

DARROW, C. W., & HEATH, L. L. Reaction tendencies relating to personality. In K. S. Lashley (Ed.), *Studies in dynamics behavior*. Chicago: Univer. Chicago Press, 1932.

DAVIS, R. C. Modification of the galvanic reflex by daily repetition of the stimulus. *J. exp. Psychol.*, 1934, **17**, 504–535.

DUFFY, E., & LACEY, C. L. Adaptation in energy mobilization in changes in general level of palmar skin conductance. *J. exp. Psychol.*, 1946, **36**, 437–452.

EDWARDS, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.

EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Appleton, 1957.

EDWARDS, A. L., & THURSTONE, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1952, **17**, 169–180.

EPSTEIN, S. Measurment of drive and conflict in humans: Theory and experiment. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962*. Lincoln: Univer. Nebraska Press, 1962.

EPSTEIN, S., & FENZ, W. D. Theory and experiment on the measurment of approach-avoidance conflict. *J. abnorm. soc. Psychol.*, 1962, **64**, 97–112.

EPSTEIN, S., & SMITH, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.*, 1956, **20**, 372–384.

EYSENCK, H. J. *Experiments in personality.* Vol. 2. London: Routledge & Kegan Paul, 1960.

FARMER, E., & CHAMBERS, E. G. Concerning the use of the psychogalvanic reflex in psychological experiments. *Brit. J. Psychol.*, 1925, **15**, 237–254.

FENZ, W. D., & EPSTEIN, S. Measurment of approach-avoidance conflict along a stimulus dimension by a thematic apperception test. *J. Pers.*, 1962, **30**, 613–632.

FREEMAN, G. L., & DARROW, C. W. Insensible perspiration and the galvanic skin reflex. *Amer. J. Phys.*, 1935, **111**, 55–63.

HATHAWAY, S. R. A comparative study of the psychogalvanic and association time measures: A new psychogalvanic apparatus. *J. appl. Psychol.*, 1929, **13**, 632–646.

HARTMANN, G. W. The facilitation effect of strong general illumination upon the discrimination of pitch and intensity differences. *J. exp. Psychol.*, 1934, **17**, 813–822.

HUNT, W. A., & LANDIS, C. Word-association, reaction time, and the magnitude of the GSR. *Amer. J. Psychol.*, 1935, **47**, 143–145.

ISTEL, J. Statistical report. *Parachutist*, 1961, **3**, 11–12.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

JUNG, C. G. *Studies in word-association: Experiments in the diagnosis of psychopathological conditions carried out at the psychiatric clinic of the University of Zürich.* New York: Moffat, Yard, 1919.

LANDIS, C. & HUNT, W. A. The conscious correlates of the galvanic skin response. *J. exp. Psychol.*, 1935, **18**, 505–529.

LANIER, L. H. An experimental study of "affective conflict." *J. Psychol.*, 1941, **11**, 199–217.

LEWIN, K. Environmental forces in child behavior and development. In C. Murchison (Ed.), *A handbook of child psychology.* Worcester: Clark Univer. Press, 1931.

LURIA, A. R. *The nature of human conflicts.* New York: Liverright, 1932.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald Press, 1944.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MILLER, N. E. Comments on theoretical models. *J. Pers.*, 1951, **20**, 82–100.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959. Pp. 196–292.

MOORE, S. C. Editorial. *Parachutist*, 1963, **4**, 5–7.

MURRAY, E. J., & BERKUN, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, **51**, 47–56.

PETERSON, F., & JUNG, C. G. Psychological investigations with the galvanometer and pneumograph in normal and insane individuals. *Brain*, 1907, **30**, 211.

RAPAPORT, D. The conceptual model of psychoanalysis. *J. Pers.*, 1951, **20**, 56–81.

RICHTER, C. P. The significance of changes in the electrical resistance of the body during sleep. *Proc. Nat. Acad. Sci.*, 1926, **12**, 214–222.

SILVERMAN, A. J., COHEN, S. I., & SHMAVONIAN, B. M. Investigation of psychophysiological relationships with skin resistance measures. *J. psychosom. Res.*, 1959, **4**, 65–87.

THORNE, F. C. The psychophysical measurment of the temporal course of visual sensitivity. *Arch. Psychol., N. Y.*, 1934, No. 170.

TREHUB, A. A theory of sensory interaction: An experimental investigation of the relationship between autonomic activity and visual sensitivity. Unpublished doctoral dissertation, Boston University, 1954.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, Rinehart & Winston, 1954.

(Received November 19, 1963)

Vol. 78, No. 9 Whole No. 586, 1964

Psychological Monographs: General and Applied

UNCERTAINTY, STRUCTURE, AND PREFERENCE[1]

HARRY MUNSINGER[2] AND WILLIAM KESSEN

Yale University

617 adults and children served as Ss in 9 studies of the relation between expressed preference and differing amounts of variability of stimulation. Random shapes and different sequential approximations to English were used as variations in stimulus variability. The results supported the following generalizations. Human beings are sensitive to amount of variability in stimulation. There is an intermediate amount of variability which was consistently most preferred by unsophisticated Ss. Preference for the stimuli used was jointly determined by number of independent characteristics of the stimuli and their meaningfulness. Preference for variability changed with Ss' experience with variable stimulation, whether the experience was induced experimentally or was the result of specific professional training. The tendency to increase preference for stimulation of high variability is related to Ss' ability to code or process variability.

A PERSISTING problem in the study of thinking has been the analysis of the circumstances under which people formulate and change the rules governing their interpretation of environmental variation. This monograph presents a series of studies designed to examine one aspect of this provoking problem. How do human beings react to variability of stimulation? What evaluation will people make of a range of stimulus events that vary widely? Specifically, we studied the preference of human subjects (*S*s) for random shapes of varying numbers of turns and their preference for sequences of letters and words that varied in sequential predictability.

Investigators of widely divergent views have considered the problems of cognitive organization that led to the research reported here. The phenomena of cognitive behavior which must be accounted for are easy enough to describe in a general way. The human response to environmental variation is neither a simple function of states of the world nor a simple function of states of the organism. In order to understand even the simplest cognitive process (or, in another language, the simplest stable hierarchy of responses to nonidentical but related stimuli), attention must be directed both to the variability of the environment and to the structure of the person (Garner, 1962; Kessen, 1962). Close recent consideration of these psychological truisms has led to the postulation of several explanatory mechanisms and the formulation of the problem in several different ways. How is "cognitive structure" developed (Piaget, 1947)? How are "plans" made (Miller, Galanter, & Pribram, 1960)? How is "competence" developed (White, 1959)? How are "informational processing systems" produced (Simon, 1962)?

Although the boundary lines are hazy, such explanatory systems and the research to which they have given rise can be classified as motivational, associationistic, or structural. Harlow (1953) and Butler (1953) have devised several ingenious demonstrations of "intrinsic motivation" and "curiosity" in higher primates. Another motivational conception of the problem of cognitive change is Berlyne's (1960). He postulates conflict as a primary mechanism, with novelty, complexity, and incongruity or surprisingness as factors which increase conflict. The reduction in conflict which follows on decreases in novelty, complexity, or incongruity is held to be reinforcing. Dember and Earl (1957) have proposed a scheme somewhat like that of Berlyne; it includes the additional postulate that discrepancy from an expected outcome is a generator of conflict. These several formulations depend chiefly on the expansion of the deprivation-satiation model of motivation to provide sources of action which go beyond physiological deficit and its repair. In an analogous expansion of a drive theory of human behavior, White (1959) has added to the libido and its derivatives an energy source that aims for mastery or "effectance."

---

[1] This research was conducted while the senior author was a United States Public Health Service Postdoctoral Research Fellow and was supported in part by a Social Science Research Council Auxiliary Research Award to the second author.

[2] Now at the University of Illinois.

Pure associationistic theories of cognitive change are uncommon. Skinner (1953) remains committed to an interpretation of response to environmental variation which directs attention almost exclusively to reinforcement contingencies. Mowrer (1960) has developed an explanation of cognitive change based on secondary reinforcement derived from second- and third-order partial-reinforcement schedules. Unlike theories of the motivational and structural sorts, associationistic models do not depend on an analysis of the relevance of a response tendency to a particular state of the environment. Such relevance must be considered in theories of "expectation," "suprise," and "schemata." Mandler (1962) has recently suggested possible lines of convergence between traditional associationistic and structural positions concerning cognitive change. He proposes a transition from associations to cognitions based on response integration and the emergence of analogic structures.

Piaget (1947) has presented what is probably the most general structural statement about cognition available. His central constructs are *assimilation*—the fitting of an environmental event to an available category or classification scheme—and *accommodation*—the development of a new category when an environmental event does not match or fit any available scheme. The interaction of these mechanisms maintains cognitive equilibrium and, granted sufficient environmental variation, results in cognitive development. In a related structural theory, Hebb (1949) has proposed a neural model of accommodation based on the formation of phase sequences. Bartlett (1932) long ago advanced the notion of "schema" and "schema with exception" as coding devices for the efficient storage of sensory input. More recently, Taylor (1960) and Newell, Shaw, and Simon (1958) have suggested that information-processing systems can handle problems of cognitive change by reference to decision theory and value systems. In general, structural theories of cognition bypass the motivational problem and assume the operation of associationistic principles; the focus of interest is on the ability of the human system to code, process, and store environmental change.

Against the variety of this background, we approached the specific problem of human preference for environmental variation and the related more general problems of cognitive organization with several fundamental assumptions. These assumptions were: (*a*) There is an apparent limitation of the ability of human beings to process environmental variation. (*b*) It is possible to escape the restriction of this limitation by means of coding rules derived from past experience with environmental variability. It is held that the interrelation of environmental variation (*stimulus variability*) and effective coding rules (*cognitive structure*) can be described as *cognitive uncertainty*. (*c*) Human beings prefer an amount of cognitive uncertainty which matches their processing ability. Cognitive uncertainty and, correspondingly, preference can change either through changes in stimulus variability or through the development of cognitive structure.

1. Limits on processing. The human system of response to environmental variation is assumed to have limits which determine amount and rate of information processing (Miller, 1956). That is, there is only a limited number of alternatives which may be handled simultaneously. Pollack and Klemmer (1954) demonstrated the limita-

tions of human processing ability by showing that listeners who make absolute judgments of pitch approach asymptotic levels of transmitted information at about 6 equally likely alternatives. Garner and Hake (1951) have shown that this limit of processing is somewhat higher for the visual system. Their *Ss* approached asymptotic limits of information processing at approximately 10 independent alternatives.

2. The effect of coding. Although the human capacity to process sensory input is limited—a limitation analogous to notions of channel capacity in communication theory (Shannon & Weaver, 1949)—immediate memory is limited only by the number of independent events presented. Miller (1956) has made this distinction in terms of "bits" for the case of transmitted information and in terms of "chunks" for the case of events which can be stored in immediate memory. Human beings can code vast amounts of information if they are chunked or structured; in this way, the apparent restrictions of the ceiling on processing ability can be escaped. For example, the span of immediate memory is in the neighborhood of seven independent items, but each of these items may contain enormous amounts of information, if variation in the environment has been subjected to a coding or chunking operation.

Coding processes that reduce cognitive uncertainty include many familiar phenomena, e.g., serial memorization and concept formation. The rest of the alphabet can be produced from memory given any single letter; it would be far more difficult to reproduce a random string of letters of the same length. Similarly, studies of concept attainment (e.g., Smoke, 1932) have shown that the use of rules that allow of efficient storage and retrieval of information markedly increase the speed and stability of learning.

3. Limit matching and preference. It is assumed further that human beings prefer a level of cognitive uncertainty which is near the limit of their ability to process. It is a matter of common observation that stimulation well below a person's capacities produces boredom and that very high levels of stimulus variation produce confusion. The present assumption extends an often heard generalization by psychologists (e.g., Davis, 1936; McClelland, 1953) and by estheticians (e.g., Ehrenzweig, 1953; Rashevsky, 1938).

These assumptions, taken together, produce a number of expectations about human preference for random shapes. The studies to be reported in this monograph were organized around the following general predictions.

1. If a wide range of stimulus variability is presented to human *Ss*, the limitation on processing ability will be seen in a nonmonotonic relation between expressed preference and variability. Specifically, an intermediate range of variability will be most highly preferred, and preference will fall off when variability is above or below this optimal level. For *random* shapes, cognitive structure ordinarily makes little contribution to the determination of cognitive uncertainty; in contrast, *stimulus variability*, defined as coordinality or number of independent turns, contributes a great deal to cognitive uncertainty.

2. The peak of the preference-variability relation—that is, the point of maximal preference—may be shifted either by changing the variability of the stimulus or by providing *Ss* with an opportunity to code stimulus variability and thereby to reduce cognitive uncertainty.

METRICS OF VARIABILITY AND PREFERENCE

The diversity of theoretical positions concerning stimulus change creates the need for a neutral metric of stimulus variability. A promising metrical system has been borrowed from communication theory (Shannon & Weaver, 1949) and extended to the measurement of environmental variation by Attneave (1959), Garner (1962), and Miller (1956), among others. This system of mathematics is based on the notion of information.

It is necessary in the specification of a stimulus array in terms of information to specify or determine the grain—the size of the smallest elements—of the array.

Specifying the effective grain of a display which is not artificially constrained is a very troublesome problem. What is needed is a measure of the information in a visual field which is not dependent upon grain. It appears that the logon content (McKay, 1956) or the coordinality of the stimulus (Attneave, 1959) may meet this need. Within broad limits, coordinality is independent of grain. However, it is almost certainly proportional to information, once a particular grain is specified. No attempt has been made in the present report explicitly to calculate number of bits of information which the stimuli contain. It is only necessary at this stage of the research to assume a positive monotonic relation between coordinality and bits of information. As a working hypothesis, based largely upon results of Pollack and Klemmer (1954), we have assumed a logarithmic relation between number of independent turns and information. Given this proportionality, the manipulation of bits of information in any stimulus display is possible, and one is not restricted to the selection of stimuli from varying numbers of alternatives as the sole means of manipulating amount of information.

The paired comparison procedure common to psychophysical research was used to provide a metric of preference. The *S*s stated their preference for one of a pair of simultaneously presented stimuli. The resulting choices were analyzed according to a procedure devised by Gulliksen (1960). Gulliksen's technique, based on a least squares solution of a complete matrix (Mosteller, 1951; Torgerson, 1958), generates scale scores of preference for each stimulus in comparison with all other stimuli in the design. These scale scores multiplied by 10 formed the metric of preference used in the studies reported here.

GENERAL METHOD

Stimulus Materials

Variations in Language. Sequences of letters and words of varying approximations to English prose and varying levels of redundancy were selected from Miller and Selfridge (1950). Two separate sets of letter and word sequences were printed on white paper in black India ink and photographed to produce white figures on black backgrounds. The photographs were mounted in slides. The sequences of letters and words are shown in Table 1.

Variations in Form (Asymmetrical). Sets of random shapes were constructed by the following method. A plywood board 100 × 100 cm. was drilled with 10,000 holes at the intersections of lines drawn 1 cm. apart. In this matrix, sets of random points were marked by placing nails in selected holes. Coordinates of the points selected were random numbers between 1 and 100 (Figure 1a). The points were then connected randomly to form a polygon. The connecting process involved two steps. First, peripheral points were connected to form a convex polygon by stretching an elastic band around the outside of the nail cluster (Figure 1b). Second, the remaining internal points were assigned letters randomly; each internal point was taken in alphabetical order into a randomly chosen segment of the surrounding polygon by hooking the appropriate segment of the elastic band over the internal nail (Figure 1c). Since lines connecting points were not permitted to cross, the number of alternative segments into which a given segment might be taken could either increase or decrease as the process continued; hence the assignment of a random sequence to the connection of the internal points. If a point fell on a straight line between two other points, a different coordinate of two random numbers was determined. This procedure generated random shapes which contained a specified number of independent discriminable turns. The shapes used in the present studies varied from 3 to 40 turns. Each shape was traced on white drafting paper, cut out, and photographed. The resulting negative, when projected, produced a black figure on a white background. This method is a specific interpretation of a general procedure for constructing random figures developed by Attneave and Arnoult (1956). The asymmetrical figures used in the studies reported here are shown in Figure 2.

Variations in Form (Symmetrical). A set of symmetrical figures was constructed in the following way. An asymmetrical polygon (constructed according to the rules outlined in the preceding paragraphs) was reflected about a vertical axis passed through the center of the 100 × 100 cm. matrix. This reflection of the shape resulted in a figure consisting of two parts, the left half of the original asymmetrical figure and its mirror image. Reflecting a figure in this way did not change the total number of turns in the shape, but did produce a substantial decrease in the number of *independent* turns (a loss of approximately 50%). The symmetrical figures used are shown in Figure 3.

The Ss

Three hundred sixteen students in introductory psychology classes and 23 art students at Yale University, 84 students from introductory psychology classes at Connecticut College, 55 students

TABLE 1

SEQUENCE OF LETTERS AND WORDS

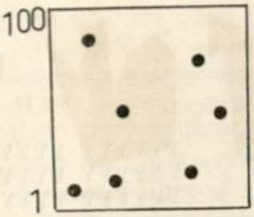
Variability of letters and words	Set I	Set II
Redundant letters	DDDDDD DDDDDD DDDDDD DDDDDD DDDDDD DDDDDD	YYYYYY YYYYYY YYYYYY YYYYYY YYYYYY YYYYYY
Redundant words	Current Current Current Current Current	Include Include Include Include Include
Prose	A short time at Alexandria is fine	He made life long friends there
Fourth-order phrases	Students always the next room are	On my rug is deep with snow
Third-order phrases	Happen to see Europe again is that	Trip to the end is coming here
Second-order phrases	Him and substance was a piano is	Is that game since he lives in school
First-order phrases	Want square chimney the wants	Especially much was said cake
Random words	Obeisance cordial dip long bed hammer	Forget lethargy fluted watch attend
Third-order words	Birs grocid ponde nome the	In no ist lat why cratict froure
Second-order words	Incore st be S deamy thall	On ei are antsoutinys T
First-order words	Tocro li rgwr nmielwis eu	Ll bnesebya tht eei the
Random letters	Ffjecyk cqsgx ydahn bixz xmrfj	Sjoml rxklr jffjuj zlp wcfckeyj

from introductory psychology classes and 19 graduate students at the University of Illinois, and 120 high school students at Amity (Connecticut) Regional School served as *S*s. The particular composition of *S*s' groups is described below for each study. No *S* was used in more than one study of preference.

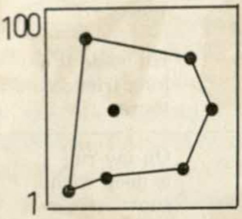
Procedure

Stimulus Display and Instructions. The stimuli were projected in pairs to a front wall of the room in which *S*s were seated. The experimenter instructed *S*s in the following words:

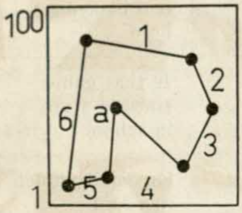
> I will show you a series of figures (sequences of letters) two at a time. I want you to indi-



1a. Illustration of the selection of random coordinates.



1b. The connection of the peripheral points.



1c. Demonstration of the connection of an internal point.

Fig. 1. Three stages in the construction of a random shape.

> cate which of the two figures (sequences of letters) you like better. You circle the *L* if you like the left figure (sequence of letters) and the *R* if you like the right one. There will be a blank slide every 10 pairs to help you keep track of the numbers. O.K.? Let's begin.

No further instructions were given concerning criteria of judgment. If questioned, *E* repeated the original instructions together with an illustration of circling the *L* or the *R*. The *S*s responded on mimeographed sheets showing numbers from 1 to 132. Opposite each number were the capital letters *L* and *R*.

Analysis. A complete paired comparison design, using sets of 8 and 12 objects, was employed in all the studies (Guilford, 1954; Gulliksen, 1960; Thurstone, 1927; Torgerson, 1958). The resulting sets of 28 and 66 item pairs were presented to groups of *S*s according to ordering procedures developed by Ross (1934). These procedures allow optimal control of time, space, fatigue, and transfer errors by randomizing side of appearance, by maximizing the distance between the same items, and by arranging the ordering of items to counterbalance fatigue and learning effects. In displaying the shapes for judgments, each pair was exposed for either 5 or 10 seconds, with a negligible interval between stimulus pairs. No difference in preference score was obtained between studies using different interval conditions so results will be described without regard to interval.

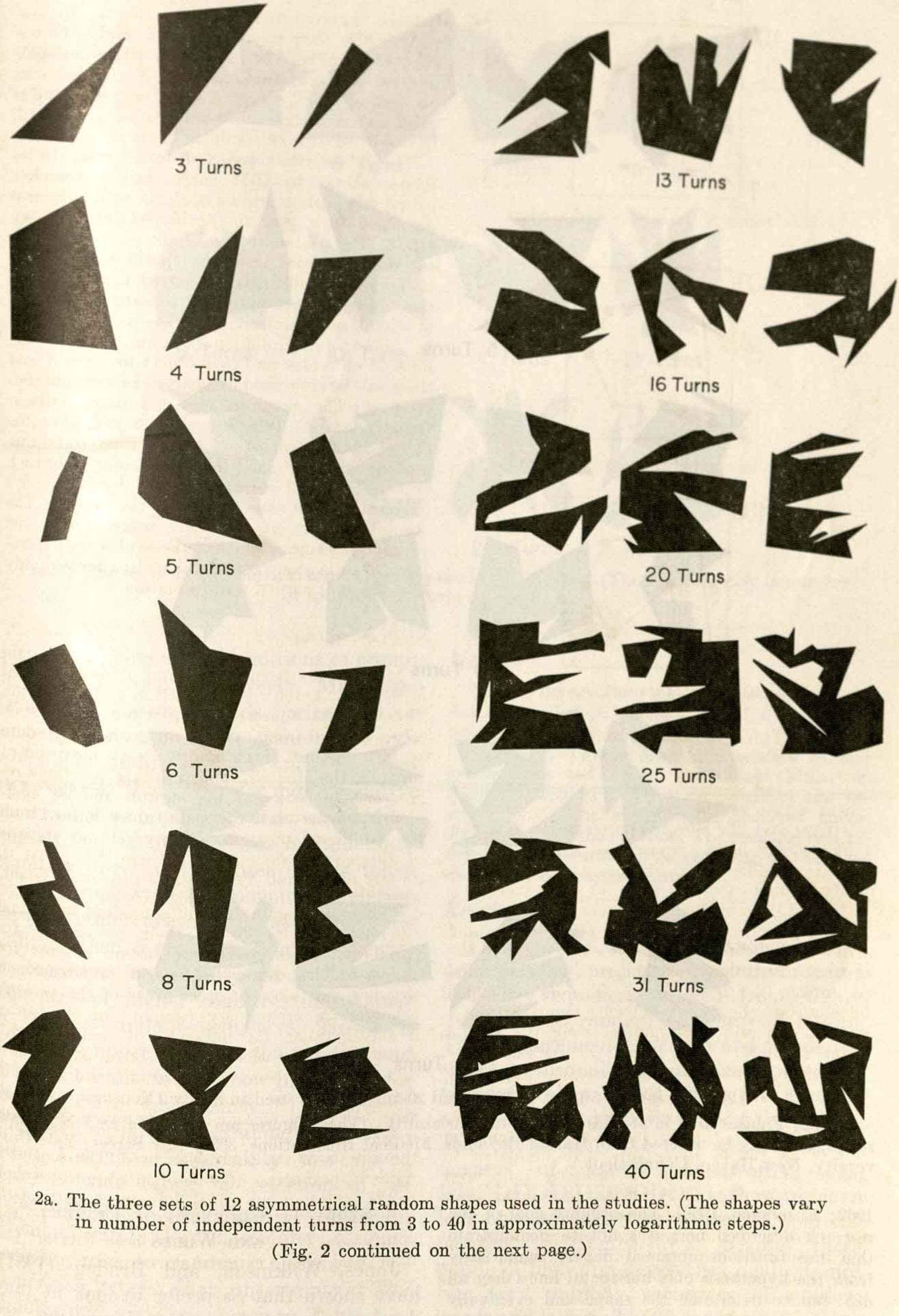
The *S*s' responses were transferred from the response sheets to IBM cards, and a complete paired comparison scaling analysis was performed according to a program developed by Gulliksen (1960). The particular case of the paired comparison design, according to the classification scheme of Torgerson (1958), is the mixed Class III with the additional assumption of treating *S*s and sets of stimuli as interchangeable. The assumption can be tested; the Gulliksen program produces the number of "votes for" each object for each *S*, and these can be compared by analysis of variance procedures.[3] The program also calculated circular triads (Kendall, 1948) for each *S* and gave the corresponding proportion, unit normal transform, arc-sine transform, and logistic transform for each object. The final scores generated by Gulliksen's program are the scale values for each object. The absolute values of the scale scores under the various transformations were little changed; therefore, all scale scores reported here are derived only from the unit normal transformation.

Statistical Note

Most results from the several studies reported in this monograph are presented in graphical form. These graphs represent the plotting of the scale scores derived from a paired comparison procedure against amount of variability (e.g., independent turns in the shapes). The resulting functions relate characteristics of the stimuli and *S*s' preference as reflected in the scale values derived from the Gulliksen program. In general, no statistic of difference is presented for these data; that is, no test of the null hypothesis is proposed (Grant,

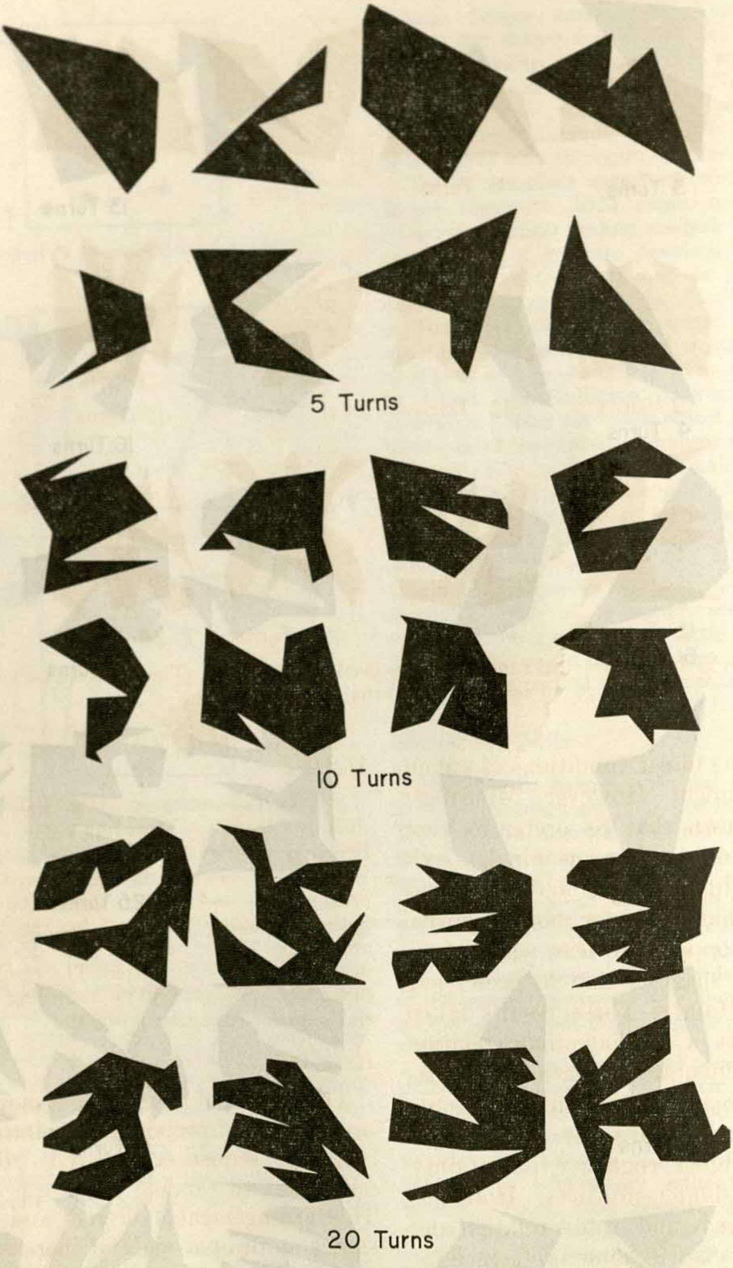
---

[3] There are significant and consistent individual differences among *S*s in their preferences for variability. It is possible to obtain a score for each *S* which indicates his level of preference and to relate this score to other areas of the person's behavior. A preliminary estimate of the correlations between this measure of preference for variability and socioeconomic status, intelligence, verbal and mathematical aptitude, and sex of *S* indicates that no simple relationships exist, although the low obtained correlations may be due to the restriction of range of these variables in the sample of the population used. The suggested index of preference may be of interest to psychologists exploring individual differences, particularly with respect to prediction of creativity. In this connection, it is interesting to note that art students and young children show similar preference scores.



2a. The three sets of 12 asymmetrical random shapes used in the studies. (The shapes vary in number of independent turns from 3 to 40 in approximately logarithmic steps.)

(Fig. 2 continued on the next page.)



2b. The sets of 5-turn, 10-turn, and 20-turn figures used in Study IX.

FIG. 2. Figures used in studies of stimulus variability. (These figures are available as 2 × 2 glass slides, and may be ordered from Harold Reynolds, Medical Illustrations, 333 Cedar Street, Yale University, New Haven, Connecticut.)

1962; Rozeboom, 1960). The primary goal of the research described here was not to demonstrate that the functions obtained depart significantly from the hypothesis of a horizontal line (they all do), but to determine the shape and eventually the parameters of the functions.

STUDY I: SEQUENCES OF LETTERS AND WORDS

Jones, Wilkinson, and Braden (1961) have shown that Ss prefer to look at randomized in contrast to alternating se-

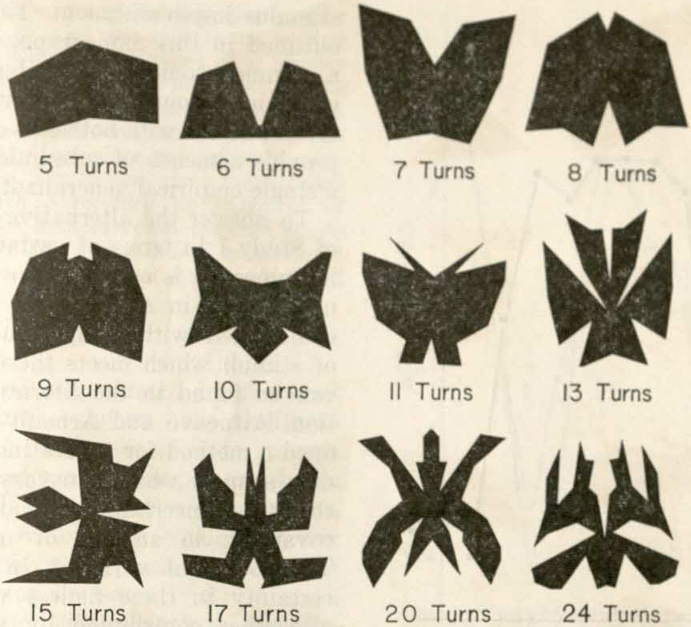


FIG. 3. The set of 12 symmetrical random shapes used in Study VII. (These shapes vary in number of independent turns from 5 to 24.)

quences of lights under conditions of stimulus impoverishment. However, Munsinger (1964) has shown that *S*s prefer to view meaningful as opposed to meaningless syllables. The finding of Jones and his associates suggests that *S*s prefer more cognitive uncertainty as opposed to less, while Munsinger's finding suggests that *S*s prefer less cognitive uncertainty. These results taken together suggest a nonmonotonic relation between preference and variations in variability. When *S* encounters small amounts of cognitive uncertainty he will prefer sequences of higher cognitive uncertainty to very redundant sequences. However, there is an upper bound to this relation and above this limit *S*s will show preference for less cognitive uncertainty. The present study was designed to examine this formulation. Cognitive uncertainty was varied across a wide range by selection of sequences of letters and words that ran from redundant letters (low uncertainty) through English prose to randomly arranged sequences of letters (high uncertainty). The stimulus materials are shown in Table 1.

Method

The S*s.* Twenty-eight male *S*s drawn from introductory psychology classes at Yale University participated in Study I.

Procedure. Twelve levels of variation in approximations to English were used (Miller & Selfridge, 1950). Procedures for selection and reproduction of the stimuli were described under the General Method section. The resulting sets of slides were presented to *S*s according to the general paired comparison procedure.

Results

The results of Study I are shown in Figure 4. The *S*s' preference is plotted against language sequences ordered from highly redundant to random sequences of letters. This arrangement of letters and words assumes an ordinal scale of increasing cognitive uncertainty from left to right. There is a decided nonmonotonic form in the results. The *S*s did not like very redundant sequences of letters and words (e.g., DDDDDDD DDDDDDD); they prefer to look at prose and close sequential approximations to English, but beyond these points of high preference they again do not like very unlikely sequences of random words and letters. The shape of the function

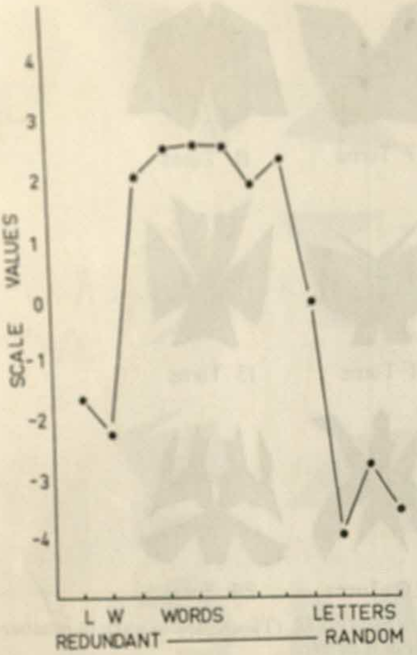


Fig. 4. Scale scores of preference for approximations to English (Study I). (Represented on the abscissa from left to right are redundant letters, redundant words, prose, fourth through first order approximations to English phrase structure, random words, third through first order approximations to English words, and random letters.)

shown in Figure 4 represents a confirmation of the postulation of a nonmonotonic relation between preference and cognitive uncertainty.

This demonstration of a nonmonotonic relation between *S*s' preference and amount of cognitive uncertainty is open to alternative interpretations. It is possible to consider preferences as a function of meaningfulness of the approximations to English. If the data shown in Figure 4 are rearranged to fit this assumption, a monotonic relation between preference and meaningfulness can be plotted. This procedure demands that very redundant and very random sequences be treated as equally meaningful, a confusing but defensible assumption. The procedure of replotting the data according to meaningfulness also ignores the finding of Jones, Wilkinson, and Braden (1961) that *S*s prefer more random sequences of lights under conditions of stimulus impoverishment. The propositions outlined in this monograph, which predict a nonmonotonic relation between preference and amount of cognitive uncertainty, are consistent with both sets of findings and provide a means of subsuming them under a single empirical generalization.

To answer the alternative interpretation of Study I in terms of variations in meaningfulness, it is necessary to vary cognitive uncertainty in such a way that it is not confounded with meaningfulness. A class of stimuli which meets these requirements can be found in the literature of perception. Attneave and Arnoult (1956) developed a method for generating stimuli (random shapes) which can vary in amount of cognitive uncertainty without necessarily covarying in amount of meaningfulness. The source of variation in cognitive uncertainty in these figures will be almost altogether contributed by what we have called stimulus variability. Sets of these shapes were constructed to test again the postulated nonmonotonic relation between preference and amount of cognitive uncertainty.

Study II: Asymmetrical Random Shapes

Three sets of asymmetrical random shapes varying in 12 approximately equal logarithmic steps from 3 to 40 turns were constructed, photographed, and presented to *S*s according to the general paired comparison procedure outlined earlier. It was expected that *S*s' preference would vary as a nonmonotonic function of variability with figures of intermediate variability preferred over figures of very high or very low variability. *Variability* was taken to be defined as the number of independent turns of the shapes (Attneave, 1959). According to this paradigm, figures of a few turns are analogous to the redundant letters and words of Study I. The wide variety in number of turns was selected to include the range of variation to which *S*s would be sensitive.

Method

The Ss. One hundred thirty-six *S*s, 44 female *S*s from Connecticut College and 92 male *S*s from Yale University participated in this study.

Procedure. Three sets of random shapes were presented to all *Ss*. The men were tested in New Haven and the women in Norwich. The display time for each pair of stimuli was 10 seconds for the male and 5 seconds for the female *Ss*. Comparable data from other samples have shown that this difference in display time does not influence preference judgments. The *Ss* were given the standard instructions. Each *S* recorded his judgment of preference on the mimeographed response sheets described earlier. The *S* circled *L* if he preferred the left stimulus of a pair and *R* if he preferred the right item.

Results

The scale scores generated by the least squares scaling solution of Gulliksen (1960) relating preference of *S*s to amount of variability of random shapes are shown in Figure 5. The two curves present data separately for males and females. As can be seen there is no appreciable difference between data for the sexes.

The postulated nonmonotonic function between variability and preference is apparent in the middle range. From 5-turn shapes, *S*s show increasing preference for greater variability until they reach a maximum at 10-turn figures. Thereafter, *S*s show a decreasing preference for figures of greater variability.[4] The three different sets of random shapes generated much the same relation between variability and preference. The correlations (eta) between sets, pooling preference judgments over *S*s, were .86, .89, and .91. Preference for figures of high variability (i.e., 25, 31, and 40 independent turns) demonstrated a departure from the expected slope. In this connection it should be noted that *S*s expressed some amusement at the many-turn figures, saying that parts of them "looked like things." It may be that the departure from prediction of the many-turn figures represent the effect of meaningfulness (Vanderplas & Garvin, 1959). As a test of this hypothesis, an independent sample of *S*s made judgments of meaningfulness in order to determine if the 25-, 31-, and 40-turn figures are more meaningful than figures in the middle range of variability (Study III).

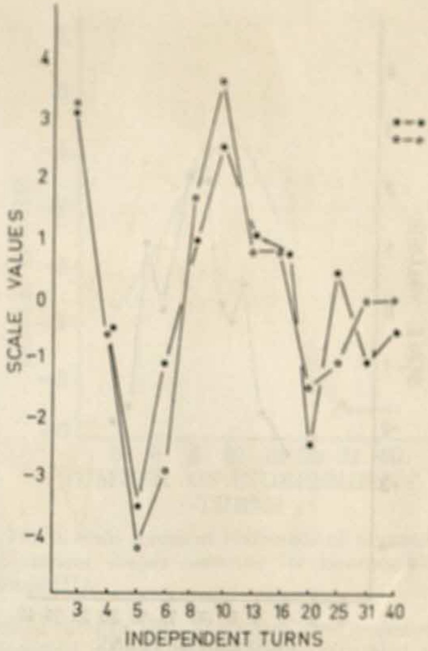


FIG. 5. Scale scores of preference for asymmetrical random shapes varying in number of independent turns from 3 to 40 (Study II). (Scores for men and women are shown.)

The unexpectedly high preference for three-turn figures is also troublesome. The effect of meaningfulness may partially account for this discrepancy from prediction and the results of Study IV are relevant to this possibility.

STUDY III: MEANINGFULNESS OF RANDOM SHAPES

Method

The Ss. Forty-eight male *S*s from introductory psychology classes at Yale University served in this study. None of the *S*s had been used in previous studies.

Procedure. Judgments of meaningfulness of the shapes of from 5 to 40 turns used in Study II were obtained. The paired comparison procedure described under general method was used with the

---

[4] The unexpectedly high preference for figures at the extremes of the array posed another problem. Exceptionally high preference for 3-turn figures may produce an artifactual reduction in preference for slightly more variable figures (i.e., 5- and 6-turn figures). Similarly, higher preference for 25-, 31-, and 40-turn figures may artifactually depress preference for slightly less variable figures. A separate study was conducted to test this interpretation. New figures were constructed in the region from 5 to 20 turns, and preferences for these determined. Happily, the results from a group of 48 *S*s took the form shown in Figure 6.

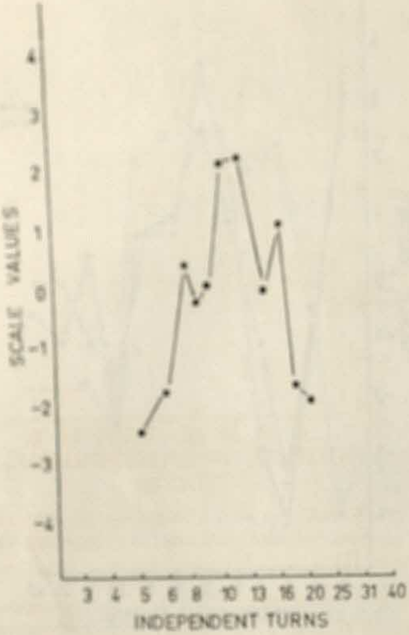


FIG. 6. Scale scores of preference for asymmetrical random shapes varying in number of independent turns from 5 to 23.

following instructions to *Ss*: "Tell me which shape reminds you of more things; that is, which shape is more meaningful." The *Ss* were told that they would see each shape for 10 seconds and that they were to make their judgments during this presentation time. The interval between presentations was negligible. Questions were answered by repeating relevant parts of the instructions.

Results

Scale scores of judged meaningfulness are presented in Figure 7. The 25-, 31-, and 40-turn figures were judged as much more meaningful than the middle range of shapes. These judgments confirm the suggestion made earlier that departures from the predicted relation between preference and amount of stimulus variability were due in part to the effects of meaningfulness. It appears that in the upper range of stimulus variability, meaningfulness of the shapes reduced the amount of cognitive uncertainty of the shapes.

The preference judgments of Study II were corrected to take account of the variation among the figures in judged meaningfulness. The correction was made by subtracting from the scale scores of preference one half the scale score of meaningfulness. The results of applying this simple subtractive correction to the data of Figure 5 are shown in Figure 8. The corrected preference scores show an excellent fit to the prediction of a nonmonotonic relation between stimulus variability and preference from five-turn figures on. As will be seen in Study VII, *ratings* by *Ss* of meaningfulness of asymmetrical random shapes generates essentially the same function as was produced by paired comparison judgments. There remains a residual deviation from the expected curve for the three-turn and the four-turn figures. This interesting deviation was the subject of Study IV.

The results of Studies II and III confirm the proposition that *Ss'* preference is a nonmonotonic function of variability. Amount of cognitive uncertainty in random shapes can be systematically changed in at least two ways. Stimulus variability (e.g., number of independent turns) can be varied and cognitive structure (e.g., meaningfulness), with its concurrent reduction in cognitive uncertainty, can be varied. In Study I (approximations to English) the major effect was probably a function of variation in cognitive structure (meaning).

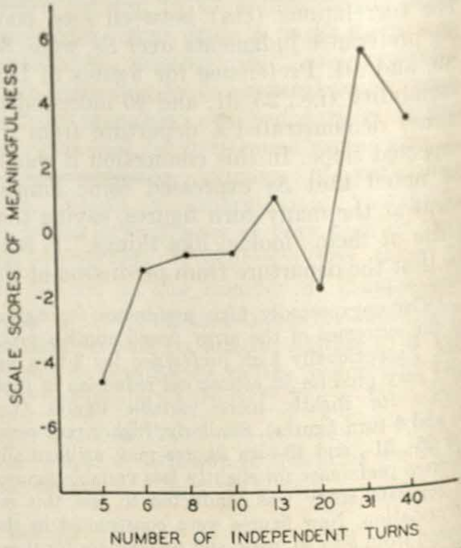


FIG. 7. Scale scores of meaningfulness of asymmetrical random shapes of varying number of independent turns (Study III).

In Study II, number of independent characteristics of the stimulus was varied and meaningfulness was relevant only because of the "projective" character of very complex figures.

Several tentative conclusions can be drawn from the first three studies in this series: (*a*) *Ss* are sensitive to variation of stimulation; (*b*) there is an intermediate amount of cognitive uncertainty which is consistently most preferred; and (*c*) preference for stimuli of varying cognitive uncertainty is jointly determined by number of independent characteristics of stimuli (variability) and meaningfulness (structure).

STUDY IV: "GOOD FORM" AND PREFERENCE

The residual deviation of preference for the three- and four-turn figures from the expected nonmonotonic function was considered in this study.

Are preferences for three- and four-turn figures deviant because of their familiarity as triangles and quadrilaterals in ordinary experience? Are the judgments deviant because such figures have "good form" in the Gestalt view (Wertheimer, 1958)? Study IV was introductory to these issues. Its chief purpose was to separate the contribution to preference judgments of variability (number of independent turns) and the "regularity" or "good form" of equilateral polygons. In addition, the role of repeated exposure in judgments of these figures was assessed by presenting the paired stimuli a number of times.

Method

The Ss. Twenty-nine *Ss* from introductory psychology classes at Yale University participated in this study.

Procedure. A set of "good form" stimuli (equilateral triangle, square, and regular pentagon) and a set of random shapes of the same number of total turns were constructed. These six figures were presented to *Ss* in a general paired comparison design. Ten complete paired comparison judgment sequences were presented to *Ss* to determine the possible effects of repeated exposure on their preference for regular or random shapes.

The presentation interval for stimuli was 5 seconds with a negligible interval between pairs. The pairs of figures were presented first to *Ss* with all shapes on a flat base. They were then presented rotated through 180°. Thereafter, each sequence of the paired comparison design was presented with the figures on a base or upside down in alternation.

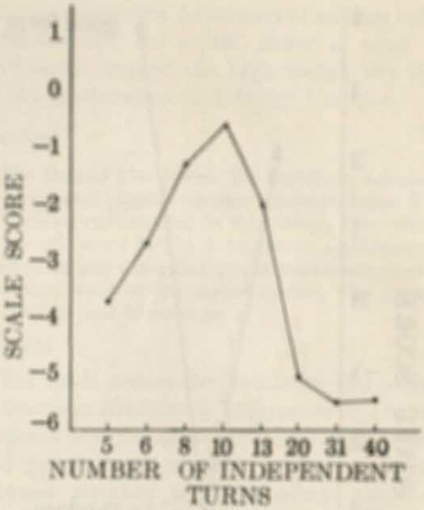


FIG. 8. Scale scores of preference of asymmetrical random shapes corrected for meaningfulness (Study III).

Instructions and response-recording techniques were identical to those used in Studies II and III.

Results

The results of the paired comparison scaling of preference for regular and random shapes of equal total turns are shown in Figure 9. Preference did not vary as a function of orientation of the figures. The preference of *Ss* for irregular and equilateral triangles was roughly equivalent as were their preferences for squares and irregular four-sided figures. The striking difference in preference between the regular pentagon and the random five-sided figure suggests the operation of a somewhat different rule from that which governs response to the three- and four-turn figures. It may be that three-sided figures are all "triangles" and four-sided figures are all "quadrilaterals," whereas in language and in simple perceptual responding *Ss* possess no easily available category inclusive of a regular pentagon and a random five-sided figure.

There was no statistically significant effect ascribable to the repeated presentation of the random and regular shapes.

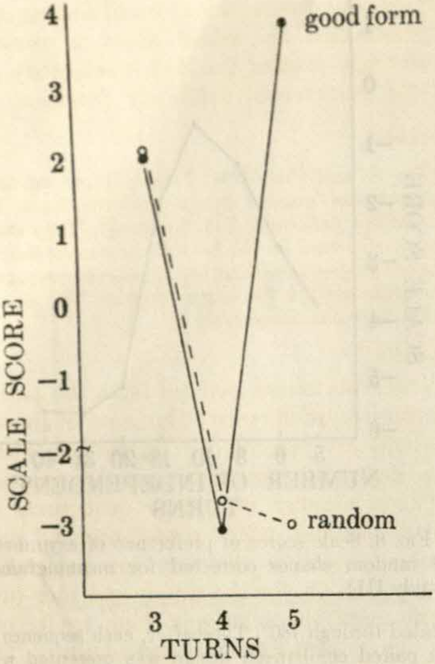


FIG. 9. Scale scores of preference for "regular" and random shapes of the same number of total turns (Study IV).

STUDY V: VERBAL FORMS, SHAPE VARIABILITY, AND RATE OF RESPONSE

The purpose of this study was to explore the relation between variation in the variability of stimuli and a nonverbal measure of preference. To assess the generality of findings based on expressed preference in the paired comparison situation, two groups of *S*s were asked to respond to materials of different levels of variability in an operant response setting. Group I was presented shapes of four levels of variability. Group II was presented four levels of variability of sequences of letters and words. It was assumed that, if expressed preference represents a general cognitive process, then nonverbal behavior should show much the same relation to variability as did stated preference.

Method

The S*s.* Thirty-six *S*s were selected and assigned to groups from a pool of *S*s in introductory psychology classes at the University of Illinois. There were 22 *S*s in the group which responded to the random shapes and 14 *S*s in the group which responded to the sequences of letters and words.

Stimulus Materials and Apparatus. Eight different 5-turn, eight different 10-turn, eight different 20-turn, and eight different 40-turn random shapes were used. Eight different forms of redundant letters, of prose, of random words, and of random letters were used in the language sequence study.

The apparatus consisted of a small black box with two windows of half-silvered mirror on the face. Stimuli were mounted on a revolving drum positioned immediately behind the mirrors. A response key was mounted on the base immediately in front of the two windows. Each time the response key was depressed, it illuminated one of the two windows and periodically activated the drum in a randomly determined variable-ratio schedule. The schedule of presentation of stimulus figures varied around a mean of 4.23:1, with a range from 1:1 to 8:1. That is, a stimulus appeared on the average after approximately four responses of the *S*.

Procedure. Within each group, all *S*s saw the eight stimuli at each of the four levels of stimulus variability. Percentage deviations from base-period performance provided scores for within-subjects analysis of variance. The sequence of presentations of levels of stimuli was randomized for each *S* with the restriction that each of the four levels of stimulation appear approximately equally often in each of the four possible positions. Within each level, the eight stimuli were presented in random order.

A 1-minute base period without stimulation preceded each of the four 5-minute periods of stimulus presentation. There was a 1-minute break after two stimulus presentation periods when stimulus tapes were changed. The *S*s were given the following instructions:

> This is a study in perception. Here is an apparatus with a key on the front and two windows. If you press the key a light will generally go on in one of the windows and you will be able to see black and white shapes [words] in the window. I want you to press the key and look at the figures [words] in the window. O.K. Any questions? Begin. There will be times when the light will not go on. Just keep pressing.

Results

The results from the two groups are presented in Figures 10 and 11. The *S*s increased their rates of responding more when presented with either 10-turn figures or prose; they do not increase their baseline rates as much when the responses produce either stimulation of low variability (5-turn figures or redundant letters) or stimulation

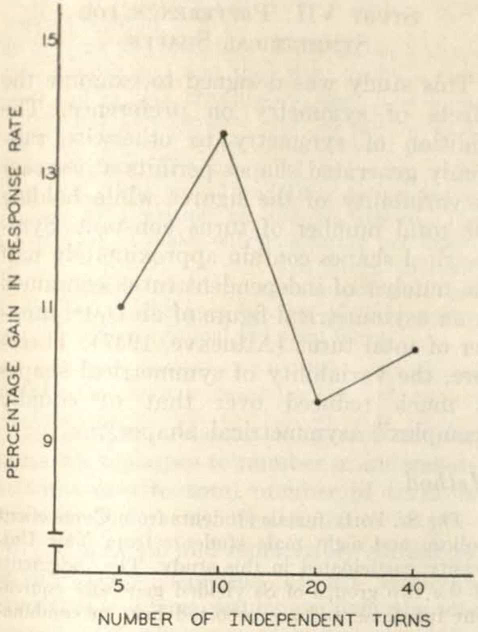


FIG. 10. Percentage increase in operant responding to four levels of random shapes (Study V).

of high variability (random words and letters; 20- and 40-turn shapes). The differences in response change among different levels of stimulus variability was significant for both random shapes ($F = 3.50$, $df = 3/36$, $p < .025$) and for language forms ($F = 7.0$, $df = 3/39$, $p < .01$). This study nicely replicates earlier findings using paired comparison designs and verbal reports. One advantage of using nonverbal procedures would seem to be the possible extension of these studies to the very young infant.

STUDY VI: THE PREFERENCE OF ART STUDENTS

The studies thus far have treated of variations in cognitive uncertainty by changing characteristics of the stimulus. It it also possible to change cognitive uncertainty by appropriate selection or training of *S*s. In the present study *S*s were selected who had experience with the invention, production, and evaluation of shapes. It was expected that, because of this experience with shapes and an assumed concurrent formation of coding rules, sophisticated *S*s would show a peak of preference toward the high-variability side of the preference-variability function.

Method

The Ss and Procedure. Twenty-three advanced painting and graphic design students from Yale Art School participated in this study. They stated their preferences for the 3- to 40-turn asymmetrical random shapes according to the paired comparison procedure used in the earlier studies. The presentation time was 10 seconds.

Results

The scale scores derived from the paired comparison preference judgments of the art students are presented in Figure 12. There is a generally monotonic positive relation between number of independent turns of the shapes and judged preferences. Of particular interest is the very low preference of art students for the three- and four-turn figures and the relatively high preference for figures of many independent turns. The art students called the three- and four-turn figures "dull," "plain," and "uninteresting."

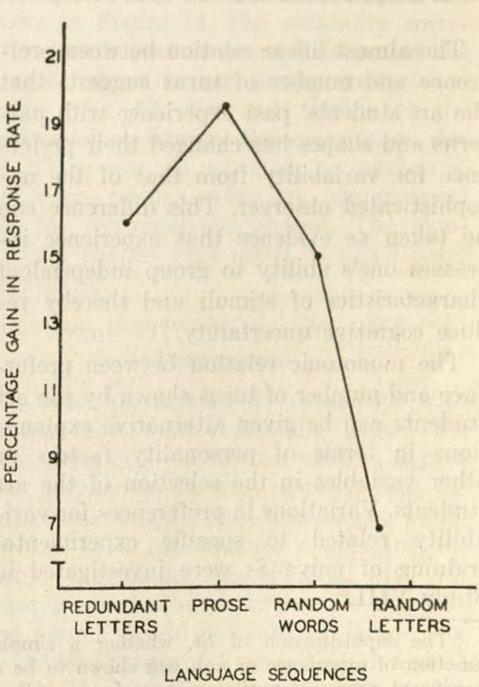


FIG. 11. Percentage increase in operant responding to four levels of variability in language (Study V).

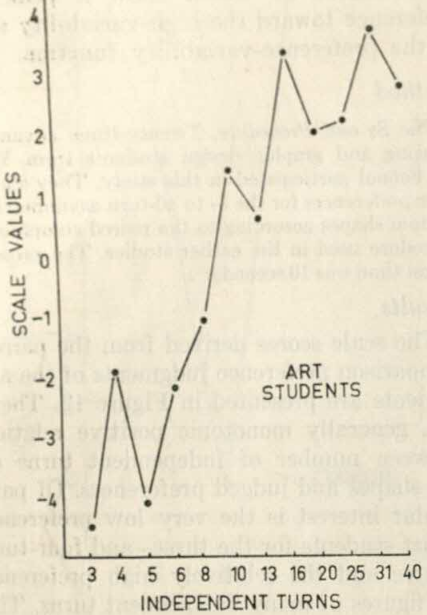


Fig. 12. Scale scores of preference of art students for asymmetrical random shapes varying in number of independent turns from 3 to 40 (Study VI).

The almost linear relation between preference and number of turns suggests that the art students' past experience with patterns and shapes has changed their preference for variability from that of the unsophisticated observer. This difference can be taken as evidence that experience increases one's ability to group independent characteristics of stimuli and thereby reduce cognitive uncertainty.

The monotonic relation between preference and number of turns shown by the art students can be given alternative explanations in terms of personality factors or other variables in the selection of the art students. Variations in preferences for variability related to specific experimental training of naive *S*s were investigated in Study VIII.[5]

Study VII: Preference for Symmetrical Shapes

This study was designed to examine the effects of symmetry on preference. The addition of symmetry to otherwise randomly generated shapes permits a decrease in variability of the figures while holding the total number of turns constant. Symmetrical shapes contain approximately half the number of independent turns contained in an asymmetrical figure of an equal number of total turns (Attneave, 1957). Therefore, the variability of symmetrical shapes is much reduced over that of equally "complex" asymmetrical shapes.

Method

The Ss. Forty female students from Connecticut College and eight male students from Yale University participated in this study. The judgments of the two groups of *S*s yielded generally equivalent functions and are reported here in combination.

Procedure. Symmetrical figures varying from 8 to 46 total turns were generated by the procedures described under General Method. These figures can be seen also as varying in number of *independent* turns from 5 to 24.

One set of 12 objects (symmetrical shapes) was presented to *S*s according to the general paired comparison procedure. Pairs of items were presented at 5-second intervals with a negligible interval between presentations. Half the *S*s saw the shapes in a position symmetrical about the horizontal plane of vision, and the other half saw the same shapes symmetrical about the vertical plane of vision. There was no difference between the curves relating preference of *S*s and number of turns under these two conditions; therefore, the data for the two kinds of orientation are presented in combination.

Meaningfulness Ratings. A separate group of 32 male *S*s rated the symmetrical shapes for meaningfulness; they also rated the asymmetrical shapes used in Study II. The *S*s were given the following instructions similar to those used by Noble (1952):

> Rate each of the figures as to the number of things or ideas it makes you think of. The ratings are made by simply writing a number opposite the space provided for that shape. The five possible ratings are described by the words *none, below average, average, above average,* and *very many.* Assign each shape a number from one to five, depending upon how many things it makes you think of. If the figure makes you think of no things, assign it a rating

---

[5] The sophistication of *S*s, whether a simple function of experience or not, was shown to be a significant source of variation in preference ($F = 12.0$, $df = 3/108$, $p < .01$) in a comparison of 19 undergraduates and 19 graduate students at the University of Illinois. The graduate students, like the art students, showed higher preference for figures of many turns.

of one. If it makes you think of many things, assign it a rating of five.

The *Ss* were told that they would see each shape for 15 seconds and that they were to make their ratings during this presentation time. The interval between presentations was negligible. Questions were answered by repeating relevant parts of the instructions. The names of the rating categories (e.g., *very many*) and their corresponding numerals were written on the blackboard. The *Ss* were told to make only one rating per figure. It was expected, by the way, that the symmetrical shapes would be rated as more meaningful than the asymmetrical shapes because of their resemblance to bilateral animal and plant forms.

Results

The function relating *Ss*' preference for symmetrical shapes to number of independent turns and to total number of turns is shown in Figure 13. The function is generally monotonic and remarkably similar to the preference function of art students (Study VI). These results confirm the expectation that cognitive uncertainty is reduced by a reduction in stimulus variability.

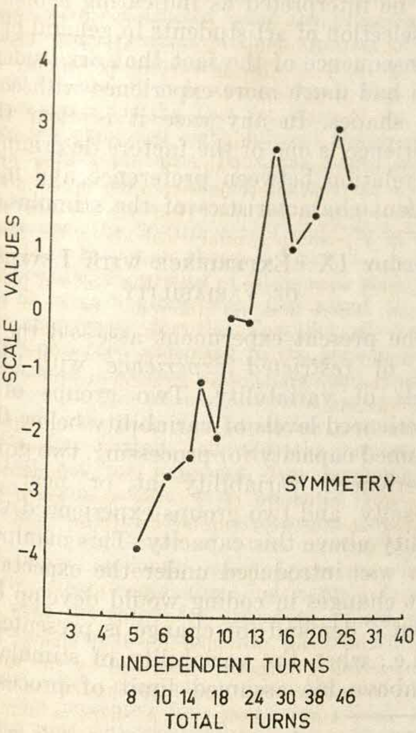


FIG. 13. Scale scores of preference for symmetrical random shapes plotted against number of independent and total number of turns (Study VII).

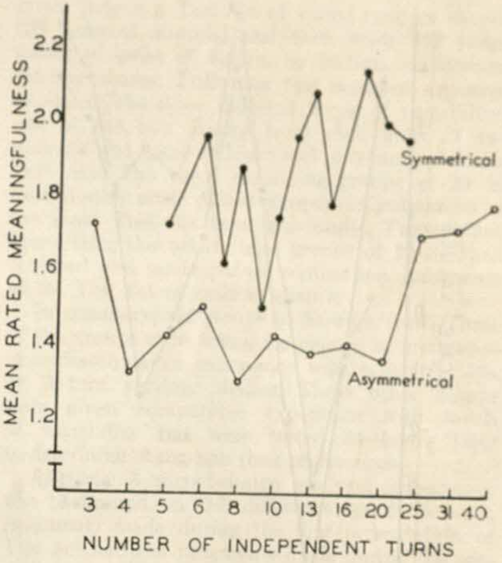


FIG. 14. Ratings of meaningfulness of symmetrical and asymmetrical random shapes (Study VII).

The ratings of meaningfulness of the symmetrical and asymmetrical shapes are shown in Figure 14. The reliability correlation (r) between two replications of the meaningfulness ratings of the symmetrical shapes was .87; for two ratings of the asymmetrical figures, the correlation was .89. There was, as expected, no overlap in the mean ratings given the symmetrical and the asymmetrical shapes. The symmetrical shapes were rated as much more meaningful.

STUDY VIII: EXPERIENCE AND PREFERENCE

In order to increase *Ss*' experience with shapes, they were repeatedly shown the same series of 3- to 40-turn random figures used previously. It was expected that *Ss*' preferences for the many-turn shapes would increase as a consequence of making repeated judgments. Specifically, with increasing experience, preference for the few-turn figures should decrease, preference for the middle range of figures should remain constant or increase, and preference for the many-turn figures should increase substantially. This prediction derives from the as-

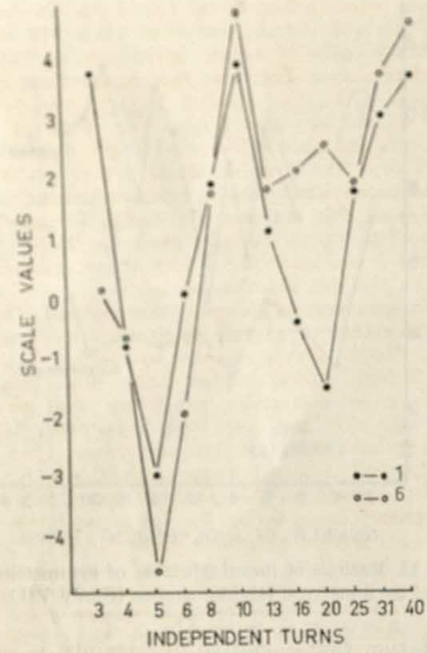


FIG. 15. Scale scores of preference for asymmetrical random shapes by *Ss* on Trials 1 and 6 (Study VIII).

sumption that *Ss* will reduce cognitive uncertainty by coding stimulus variation and thereby increase the amount of variation in the environment which can be processed.

Method

The Ss. Thirty-one male *Ss* from introductory psychology classes at Yale University served in this study.

Procedure. Twelve 3- to 40-turn asymmetrical random shapes were presented to *Ss* in six sequences of the paired comparison design. Thus, each *S* judged a total of 396 pairs of shapes in the course of a 2-hour experimental session. Each pair of shapes was presented for 10 seconds with a negligible interval between presentations.

Results

The results of the judgments of *Ss* over the first and last trials of the six repetitions of the paired comparison design are shown in Figure 15. As predicted, the curves show a shift toward reduced preference for the shapes of few turns, little change in the middle range, and a substantial increase in preference for the many-turn figures. An $S \times$ Condition $\times$ Repeated Measures analysis of variance was performed on these data. To insure unconstrained row variance, number of "votes for" the 3-, 5-, 10-, 13-, 20-, and 40-turn figures were used as scores. The main treatment effect (number of independent turns) was significant ($F = 51$, $df = 5/135$, $p < .001$) as was the interaction between *Ss* and treatments ($F = 2.2$, $df = 135/135$, $p < .01$). The prediction that experience with the figures changes the shape of preference was confirmed by a significant interaction between number of independent turns (treatments) and trials ($F = 8.8$, $df = 5/135$, $p < .001$).[6] These results demonstrate that *Ss* shift in preferences for variability as a result of experience with the figures and that different *Ss* prefer different amounts of cognitive uncertainty. However, even after six repetitions of the paired comparison design, the unsophisticated *Ss* do not show the monotonic pattern of preference which the art students showed in Study VI. This finding may be interpreted as indicating a bias in the selection of art students in general or as a consequence of the fact that art students have had much more experience with complex shapes. In any case it is clear that experience is one of the factors determining the relation between preference and independent characteristics of the stimulus.

STUDY IX: EXPERIENCE WITH LEVELS OF VARIABILITY

The present experiment assessed the effect of *restricted experience* with three levels of variability. Two groups of *Ss* experienced levels of variability below their assumed capacity for processing, two groups experienced variability at or near this capacity, and two groups experienced variability above this capacity. This manipulation was introduced under the expectation that changes in coding would develop only when a demand for change is presented to *S*, i.e., when the variability of stimulation is above his assumed limit of processing ability. Two groups were run at each level of experience in order to determine whether the expected change in behavior following experience with variability is a short-term (adaptation) effect or a long-term (learning) effect. One group of each pair was tested immediately after experience with a level of stimulus variability; the other group in each pair was tested 24 hours after such experience.

---

[6] The failure to show a significant shift in preference with experience of few-turn figures (Study IV) may be attributed to the fact that all of the figures in Study IV lay below *Ss'* assumed limit of processing ability.

Change in expressed preference for random shapes was used as a measure of the effect of experience with variability. It was predicted that only experience with a level of variability above the assumed limit of processing ability would result in a significant increase of preference for figures of high variability.

Method

The Ss. One hundred twenty students from Amity Regional High School served as *Ss* in this experiment.[7]

Stimulus Materials. Four sets of eight random shapes were constructed using the procedure described under General Method. One set of eight random shapes contained figures of 5, 6, 8, 10, 13, 20, 31, and 40 independent turns. This set of shapes was called the Test Set. Each of the other three sets contained eight different shapes of the *same* number of turns. Thus, the 5-turn set was made up of eight different figures of 5 independent turns, the 10-turn set of eight 10-turn figures, and the 20-turn set of eight 20-turn figures. These shapes are shown in Figure 2.

Procedure. Six groups of 20 *Ss* each were used. The *Ss* in each group were first asked to state their preferences for the Test Set of random shapes presented according to the general paired comparisons procedure with a 10-second presentation interval. This allowed an estimation of a base-level of preference for variability for all the groups. Immediately following the presentation of Test Set for judgment, two randomly selected groups of *Ss* were presented four successive, complete, paired comparison preference runs (112 judgments) of 5-turn random shapes. Two other randomly selected groups, immediately after completing judgments of Test Set, were presented for successive paired comparison preference runs with the 10-turn random shapes. The remaining two groups, immediately after completing judgments of Test Set, were presented four successive paired comparison runs with the 20-turn shapes. Thus each group judged a Test Set of varied random shapes (28 pairs of stimuli) and then made 112 judgments of pairs of 5-turn, or 10-turn, or 20-turn random shapes. Following this repeated exposure to one of the three different levels of variability, one of the two groups from each level of experience was asked to leave and to return again the next day. The three remaining groups of 20 *Ss* immediately made paired comparison judgments of the same Test Set seen previously. Twenty-four hours later the other three groups of 20 *Ss* each returned and made paired comparison judgments of the Test Set of random shapes.

In summary, six groups of *Ss* were used. Three of the groups were tested for change in preference immediately after experience with either 5-, 10-, or 20-turn random shapes. Three other groups were given comparable experience with levels of variability but were tested 24 hours later to determine changes in their preferences.

Analysis. A mixed-design analysis of variance was performed on the difference scores between judgments made during the first presentation of Test Set and the judgments made during the second presentation of Test Set. The response measure used to compute the difference score was the number of "votes for" each random shape in Test Set by each *S*. To insure unconstrained row variance only the votes for the 5-, 10-, 20-, and 40-turn shapes were used.

Results

The major prediction of the experiment—that experience with variability *above* the assumed limit of processing ability will result in an increase in preference for figures of high variability—can be assessed by examining the difference scores based on the shift in preference from the first presentation of Test Set to the second presentation of Test Set. These data are shown in Figure 16. The *Ss* who make many judgments of 10-turn shapes show *increased* preference for 10-turn shapes and *decreased* preference for 20-turn shapes when tested 24 hours after experience with the shapes. The relevant statistic of assessment for this finding is an interaction among levels of experience, shapes in the Test Set, and time of testing ($F = 3.95$, $df = 6/342$, $p < .01$). When high school *Ss* are exposed to stimulus variability at or near their processing limit, they do not develop new coding procedures which permit them to handle (prefer) more variability. On the contrary, the data of Figure 16 indicate that, under such conditions of experience manipulation, there is an exag-

---

[7] The authors wish to express their appreciation for the cooperation of J. Foley and his staff of the Amity Regional High School.

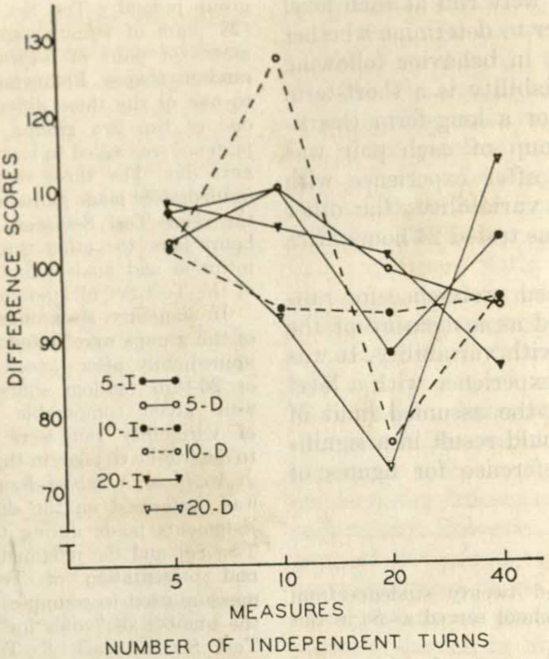


Fig. 16. Difference scores between test trial 1 and test trial 2 for the six groups of 20 *S*s, tested immediately (I) or after 24 hours (D), plotted over the four levels of measurement of variability. (Study IX).

geration of the "normal" preference for figures of 10 turns with a concurrent reduction in preference for figures of higher variability 24 hours after exposure. The *S*s exposed to stimulus variability well above their processing limit, on the other hand, appear to develop coding procedures which allow them to handle (prefer) figures of higher variability with a concurrent reduction in preference for figures of lower variability 24 hours after exposure. It is interesting that these striking effects show up only after a 24-hour interval; clearly, the effects shown are not the result of adaptation or stimulus satiation.

There were no other statistically significant changes in the preference of *S*s for figures from the first to the second presentation of Test Set.

Discussion

Several generalizations may be drawn from the series of studies presented here. Human adults are sensitive to differences in variability of stimulation. There is an intermediate amount of cognitive uncertainty which is consistently most preferred. Cognitive uncertainty is jointly determined by number of independent characteristics of the stimuli and their rated meaningfulness. Preference varies with *S*s' experience of variability, whether the experience is induced experimentally or is the result of specific professional training. Moreover, for high school *S*s, experience with variability above their assumed limit of processing ability results in an increase in preference for figures of high variability.

These results can be fit easily to an explanation that has as its central construct *cognitive uncertainty*. Human beings are able to handle only a limited amount of cognitive uncertainty. Cognitive uncertainty, an interrelation of environmental variability and cognitive structure, can be modified by changes in either of its components. Variability, assumed to be a characteristic of stimuli, can be varied by changes in the redundancy of stimulation, e.g., symmetry or duplication of components. Cognitive structure, assumed to be a characteristic of the person, can be changed by providing the person with an opportunity to code or chunk stimulus variability through experience. Further, a person will most prefer environmental input which is at or near the limit of his ability to handle cognitive uncertainty.

This simple model provides a useful approach to the class of problems usually considered under such titles as "curiosity," "exploratory behavior," and "reactions to novelty and complexity." For example, some treatments of novelty have suggested that any class of unfamiliar stimuli will be preferred over a class of familiar stimuli. Under the present formulation, novelty is a special case of cognitive uncertainty—a compound of the variability of stimulus events and of a person's ability to code variability. A stimulus will be treated as "novel" (and *S*s will prefer it) as long as it contains an amount of cognitive uncertainty just above his capacity. A stimulus containing markedly less uncertainty will not be treated as novel unless *S* is in a state of stimulus impoverishment. A stimulus containing markedly more cognitive uncertainty than *S*'s capacity, however novel it may be when defined against his experience, will not be preferred to less novel but capacity-matching stimulation.

Expectations about curiosity in the present formulation depend on the assumption of maximal preference for cognitive uncertainty near the person's processing limits. The organism will, *ceteris paribus*, ignore stimulation below the limit, tend to approach ("curiously") stimulation at or just above this limit, and ignore or turn away from stimulation well above the limit of processing ability.

The theoretical formulation sketched here has obvious developmental implications, and it generates explicit predictions concerning the preference of children for varying amounts of cognitive uncertainty. In those cases where no ordinary opportunity has been provided for chunking or structuring during the course of growing up (e.g., the case of random shapes), it is predicted that there will be no developmental differences in peak of preference of the preference-variability function. That is, the postulated limited capacity for handling cognitive uncertainty will not produce a difference between children and adults. The experience that increases with age can influence this limitation only by providing procedures for coding variability. Thus, unlike the case of random shapes, it is predicted that clear developmental differences will appear in the study of preference for language sequences. In this instance, the child's increasing experience with the forms of English can be seen as a way of increasing his ability to code language sequences; and the experience should result in preference for more and more variability of language as the child grows. These predictions have been confirmed in a recent study (Munsinger, Kessen, & Kessen, 1964).

Among the most intriguing implications of this model of limited capacity and preference is that the matching of cognitive uncertainty to capacity is an important mechanism for behavioral change. By assumption, this matching may take place by increasing variability from levels too low or by decreasing it from levels too high. This simple postulate then takes account both of the perplexing tendency of most animals to "seek stimulation" and of the "self-reinforcing" properties of exploration. A word more on this last characteristic is in order. It is expected that the organism will approach stimulation just above his capacity. If he codes some of the resident variability by whatever technique, his apparent capacity is thereby raised and he can handle still another increase in variability. In this way, oftentimes without intervention from other organisms, the "exploring" animal or person will increase his knowledge of the environment and his cognitive competence. Studies aimed at the clarification of these issues are now in progress.

SUMMARY

Nine studies of human preference for environmental variation were carried out under the following assumptions. (*a*) There is an apparent limitation of the ability of human beings to process environmental variation. (*b*) It is possible to escape the restriction of this limitation by means of

coding rules derived from past experience with environmental variability. It is held that the interrelation of environmental variation (stimulus variability) and effective coding rules (cognitive structure) can be described as cognitive uncertainty. (*c*) Human beings prefer an amount of cognitive uncertainty which matches their processing ability. Cognitive uncertainty and, correspondingly, preference can change either through changes in stimulus variability or through the development of cognitive structure.

In Study I ($N = 28$) *S*s made paired comparison judgments of preference for language sequences varying from highly redundant to random. It was found that they most preferred stimuli of intermediate variability. To reduce the contribution of experience to judgments of preference for stimulus variability, random shapes were constructed according to the techniques of Attneave and Arnoult and were used in Study II ($N = 136$). The *S*s made paired comparison judgments of preference for random shapes of from 3 to 40 independent turns. As expected, preference was greatest for figures of an intermediate number of turns. There was an unexpectedly high preference for figures of very few (3 and 4) and very many (25, 31, and 40) turns. Study III ($N = 48$), in which *S*s made paired comparison judgments of meaningfulness of random shapes, showed that the high preference shown for figures of many turns is accounted for by their relatively higher meaningfulness. Studies IV and VII ($N = 29$) showed that no simple hypothesis of meaningfulness can account for preference for 3- and 4-turn figures. In Study V ($N = 36$) key-depression which produced the stimuli was the response measure in a partial replication of Studies I and II. As expected, the nonmonotonic relation between stimulus variability and response was found again, with materials of intermediate stimulus variability showing the greatest effect on behavior.

Having shown the general character of the relation between preference and stimulus variability, we turned in the last four studies to an examination of the ways in which cognitive uncertainty and, thereby, preference could be modified by changes in either stimulus variability or cognitive structure. Twenty-three art students participated in Study VI, judging random shapes of the kind used in earlier studies. Unlike naive *S*s, art students generate a monotonic increasing relation between stimulus variability and preference. In Study VII ($N = 48$), a similar monotonic function was shown, not by selection of sophisticated *S*s, but by varying the stimulus variability (specifically, symmetry) of the figures judged by naive *S*s. In Study VIII ($N = 31$), an attempt was made to build coding rules for the random shapes by presenting them for judgment a large number of times. As predicted, *S*s showed, after experience with the figures, an increased preference for figures of many turns. Study IX ($N = 120$) subjected *S*s to limited experience with the shapes (i.e., only 5-, 10-, or 20-turn figures) and showed that degree and direction of change in preference is related to the nature of the intervening experience and to time of testing the change. Eighty-six *S*s were run in two studies of method to rule out alternative interpretations of the data.

The following conclusions were drawn from these studies. Human adults are sensitive to differences in variability of stimulation. There is an intermediate amount of cognitive uncertainty which is consistently most preferred. Cognitive uncertainty is jointly determined by number of independent characteristics of the stimuli and their judged meaningfulness. Preference varies with *S*s' experience of variability, whether experience is induced experimentally or is the result of specific professional training. Interpretation of these results in terms of an interrelation of stimulus variability and cognitive structure provides a useful approach to problems of curiosity, exploratory behavior, and reactions to novelty, as well as providing a suggestive model for developmental change.

REFERENCES

Attneave, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53,** 221–227.

Attneave, F. *Applications of information theory to psychology*. New York: Holt, 1959.

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53,** 452–471.

Bartlett, F. C. *Remembering: An experimental and social study*. London: Cambridge Univer. Press, 1932.

Berlyne, D. E. *Conflict, arousal, and curiosity*. New York: McGraw-Hill, 1960.

Butler, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46,** 95–98.

Davis, R. C. An evaluation and test of Birkhoff's aesthetic measure formula. *J. gen. Psychol.*, 1936, **15,** 231–240.

Dember, W. N., & Earl, R. W. Analysis of exploratory, manipulatory and curiosity behaviors. *Psychol. Rev.*, 1957, **64,** 91–96.

Ehrenzweig, A. *Psychoanalysis of artistic vision and hearing*. London: Routledge & Kegan Paul, 1953.

Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58,** 446–459.

Grant, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1962, **69,** 54–61.

Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

Gulliksen, H. An IBM program for a complete paired comparisons schedule. Technical Report, 1960, Project NR 150-088, Nonr 1858-(15).

Harlow, H. F. Mice, monkeys, men, and motives. *Psychol. Rev.*, 1953, **60,** 23–32.

Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.

Jones, A., Wilkinson, H. J., & Braden, I. Information deprivation as a motivational variable. *J. exp. Psychol.*, 1961, **62,** 126–137.

Kendall, M. G. *Rank correlation methods*. London: Griffin, 1948.

Kessen, W. "Stage" and "structure" in the study of children. In W. Kessen, & C. Kuhlman (Eds.), Thought in the young child. *Monogr. Soc. Res. Child Develpm.*, 1962, **27** (No. 2, Whole No. 83).

McClelland, D. A. *The achievement motive*. New York: Appleton-Century, 1953.

McKay, D. M. Towards an information-flow model of human behavior. *Brit. J. Psychol.*, 1956, **47,** 30–43.

Mandler, G. From association to structure. *Psychol. Rev.*, 1962, **69,** 415–427.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63,** 81–97.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior*. New York: Holt, 1960.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63,** 176–185.

Mosteller, F. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 1951, **16,** 3–11.

Mowrer, O. H. *Learning theory and the symbolic processes*. New York: Wiley, 1960.

Munsinger, H. L. Meaningful symbols as reinforcing stimuli. *J. abnorm. soc. Psychol.*, 1964, **68,** 665–668.

Munsinger, H., Kessen, W., & Kessen, M. L. Age and uncertainty: Developmental variations in preference for variability. *J. exp. child Psychol.*, 1964, **1,** 1–15.

Newell, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem solving. *Psychol. Rev.*, 1958, **65,** 151–166.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **53,** 421–430.

Piaget, J. *The psychology of intelligence*. London: Routledge & Kegan Paul, 1947.

Pollack, I., & Klemmer, E. T. The assimilation of visual information from linear dot patterns. Technical Report No. 54-16, 1954, Air Force Cambridge Research Center.

Rashevsky, N. Contribution to the mathematical biophysics of visual perception with special reference to the theory of aesthetic values of geometrical patterns. *Psychometrika*, 1938, **3,** 253–271.

Ross, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. *J. educ. Psychol.*, 1934, **25,** 375–382.

Rozeboom, W. W. The fallacy of the null hypothesis significance test. *Psychol. Bull.*, 1960, **57,** 416–428.

Shannon, C. E., & Weaver, W. *The mathematical theory of communications*. Urbana: Univer. Illinois Press, 1949.

Simon, H. A. An information processing theory of intellectual development. In W. Kessen, & C. Kuhlman (Eds.), Thought in the young child. *Monogr. Soc. Res. Child Developm.*, 1962, **27** (No. 2, Whole No. 83).

Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.

Smoke, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**(No. 4, Whole No. 191).

Taylor, D. W. Toward an information processing theory of motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1960*. Lincoln: Univer. Nebraska Press, 1960.

THURSTONE, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

VANDERPLAS, J. M., & GARVIN, E. A. The association of random shapes. *J. exp. Psychol.*, 1959, **57**, 147–154.

WERTHEIMER, M. Principles of perception. In D. C. Beardslee, & M. Wertheimer (Ed.), *Readings in perception*. New York: Van Nostrand, 1958.

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychol. Rev.*, 1959, **66**, 297–333.

(Received August 15, 1963)

Vol. 78, No. 10 Whole No. 587, 1964

Psychological Monographs: General and Applied

PHYSIOLOGICAL ACTIVITY AND PERFORMANCE:

A COMPARISON OF KINDERGARTEN CHILDREN WITH YOUNG ADULTS[1]

ROGERS ELLIOTT[2]

Allan Memorial Institute, McGill University

21 young adults and 41 kindergarten children performed repeated blocks of trials of simple auditory reaction time (RT), sometimes for incentives, sometimes not, while EEG and other physiological responses were recorded. Children differed from adults in that they manifested (a) no covariation between quality of performance and level of physiological activity; (b) far lower intra-individual correlations between one physiological response and another; (c) no adaptation over the session; (d) far weaker relations between preparatory (interstimulus) intervals and RT; and (e) increases, rather than decreases, in the amplitudes of the various EEG frequencies with increasing motivation and physiological activity. These differences were considered in the light of certain important similarities in response between the children and various groups of adults.

THIS investigation was undertaken primarily to examine two relations in children which have heretofore been studied only in adults. The first of these relations is between muscular and autonomic activity, on the one hand, and quality of perceptual-motor performance, on the other. The second is the relation between EEG phenomena and motivation.

Data pertaining to the nature of the relation between physiological activity and efficiency of performance have been comprehensively reviewed by Duffy (1962). In most studies of adults some relation has been reported between one or more physiological indicators and quality of performance. Duffy interprets such relations, in the main, as supportive of an activation theory of performance; i.e., quality of performance is best when physiological activity is neither too high nor too low, but is at some optimal moderate level. But any theoretical interpretation of such literature is certainly limited in terms of age: There have been no studies, with children as subjects (*S*s), in which one or more physiological responses has been measured during reasonably sustained perceptual-motor performance. This gap in knowledge prompted the present study.

There have been some well-known normative and developmental studies of EEG (Corbin & Bickford, 1955; Henry, 1944; Lindsley, 1939; Walter, 1950). But as Walter (1950) pointed out, studies of EEG have been done primarily while *S*s were at rest, or hyperventilating, or opening and closing their eyes—activities of somewhat limited psychological interest. No study has examined EEG characteristics of children while they sustain performance at some task, nor have such EEG characteristics been systematically related either to motivational variables or other physiological variables. This study was designed to make possible such an examination of EEG in children. In addition, it proved possible, owing to techniques introduced by MacNeilage (1961), to examine very brief fluctuations of different EEG frequency components during the sequential phases of a simple, stimulus-response (S-R) occurrence.

---

[1] This research was conducted while the author was a United States Public Health Postdoctoral Research Fellow in the Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, McGill University. The research was aided by a contract with the Surgeon General's Office of the United States Army; and by grants from the National Institute of Mental Health (United States Public Health Service), Department of National Health and Welfare (Canada), National Research Council of Canada, and the Defence Research Board, Canada. R. B. Malmo sponsored the research, and his very generous support is gratefully acknowledged, as is the contribution of W. Mundl, who designed much of the equipment and helped in its operation. P. MacNeilage, R. B. Malmo, and J. Weiffenbach assisted the author materially with extensive criticism of the manuscript. The author wishes particularly to express his gratitude for the cooperation of the parents and children involved.

[2] Now at Dartmouth College.

Given these major areas of exploration, a design modeled fairly closely to a study of adults by Stennett (1957b) was constructed. That is to say, a perceptual-motor task was to be performed under conditions in which incentive level was varied; and a variety of physiological responses, including EEG, were to be simultaneously and continuously recorded. Kindergarten children were employed as *S*s, particularly because they seemed to be near the lower age limit at which they could endure the rigors of extensive physiological measurement, and because they could maintain interest in a task for a reasonable period of time. Adults were used as a direct-comparison group and as a check on the representativeness of the design—i.e., a check on the expectation that results based upon the present sample of adults ought to be similar to results obtained in other studies of physiological activity and quality of performance in adults. A task was chosen which was well within the capacity of the children, namely, simple serial auditory reaction time (RT).

Apart from its utility in an exploratory study of relations in children among autonomic, muscular, EEG, and performance variables, the design generates data which permit analyses bearing informatively on two additional areas of current interest in human psychophysiology. These areas have to do with physiological concordance and physiological adaptation.

Physiological concordance refers to the covariation over time and occasions of different physiological responses within the individual. Reports from adult samples (Hovland & Reisen, 1940; Lazarus, Speisman, & Mordkoff, 1963; Schnore, 1959) have provided evidence for substantial degrees of such concordance; and such evidence has been interpreted as supporting the existence of a physiologically based communal drive mechanism, or general activation system (Duffy, 1962; Lazarus et al., 1963; Malmo, 1959). On the other hand, Lacey (1959) has presented data which imply strongly that the degree of positive covariation of any two physiological responses is reduced as a function of the number of different tasks or situations in which *S* is required to participate. Thus, concordance would be higher with repetitions of a single task than with presentation of a variety of tasks. The power of a given situation to produce its own *pattern* of physiological responses (rather than a general *level* of response) has been emphasized by Davis (1957), Ax (1953), Schacter (1957), and Wolf and Wolff (1943); and Lacey (1959) has termed the phenomenon "situational stereotypy." As one might by now surmise, data on concordance from children are lacking.

Physiological adaptation, or habituation, refers to a decline in the activity level of a response over time. The classic illustration is from Duffy and Lacey (1946), showing the decrease in skin conductance within parts of a session, within complete sessions, and from session to session as college girls repeated auditory discrimination tasks. As with concordance, adaptation is greatest with repetitions of a single task and is reduced by the presentation of different tasks or different levels of difficulty of a given task (for a general discussion, see Woodworth & Schlosberg, 1954, Ch. 6). Another factor known to affect the adaptation of physiological responses is the clinical status of *S*. Compared with normals, psychoneurotics adapt less in muscle-tension (EMG) responses to repeated heat-pain stimuli (Malmo & Wallerstein, 1955) and to repeated tones (Bartoshuk, 1959). Schizophrenics adapt less in heart rate (HR) during repeated heat-pain stimulation (Cohen & Patterson, 1937) and less in GSR during repeated auditory stimulation (Shakow, 1962). Habituation to repeated auditory stimuli has been reported in the HR of infants (Bartoshuk, 1962) and in the GSR of groups of 4- (Jones, 1935) and 7-year-old children (Corah, 1963). But adaptation by children to more complex tasks and stimuli, such as those used by Duffy and Lacey (1946) or Malmo and Wallerstein (1955), has not been investigated; nor have children been compared directly with adults with reference to adaptation.

One final area of analysis was possible owing to the nature of the RT task. The trains of tones used followed each other at varying intervals in a manner somewhat like that used by Telford (1931). It was possible to consider these varying interstimulus intervals as preparatory intervals and to examine the effect of such intervals upon reaction speed as a function of age. Much has been made of the relation between preparatory interval and RT with reference to both schizophrenia (Shakow, 1962), and to the aging process (Botwinick, Brinley, & Birren, 1957), and it was thought useful to take at least a cursory look at data from children in this case.

In summary, what is being presented is a highly exploratory, empirical study whose primary aim is to get information on the relations between physiological activity and quality of performance and on EEG phenomena in children. In addition, data concerning physiological adaptation, physiological concordance, and the effect of preparatory interval on RT are to be developed and discussed. There were little if any data, to say nothing of developed theory, to guide predictions in any of these areas, and none were made.

METHOD

The Ss

The *S*s were 41 kindergarten children (22 boys and 19 girls) whose average age was 6.0 years (range: 4.9–6.6), and 21 adults (11 males, 10 females) whose average age was 25.5 (range: 17–38). There was no significant difference in age between males and females in either age group. Thirty-four of the 41 children were selected from the morning and afternoon kindergarten school classes in a middle-class residential district of Montreal.[3] These 34 were a subgroup of the 60 whose parents were sent a letter asking them to volunteer for the study. (The parents of 15 refused.) The other 7 children were obtained through acquaintance with members of the medical and research staffs of the Allan Memorial Institute of Psychiatry. This sample of children is biased toward the higher levels of socioeconomic background. The most frequent occupation of the fathers was engineer; the next most frequent was salesman; and there were from two to four fathers in each of the categories doctor, artist, journalist, teacher, and executive. Cases in which the head of the family was a clerical or blue-collar worker were infrequent. Only two children came from broken homes.

A greater bias in the sample resulted from deliberate selection by the experimenter (*E*). Letters asking for volunteers were sent only to the parents of children who, on the basis of teachers' ratings, appeared neither too sensitive ("cries frequently") nor too active ("is never quiet") for the requirements of the experiment. Thus, the parents of 15 of the 75 children were never even approached. Despite this precaution 9 children failed to complete the first session, and 2 others could not be persuaded to return for the second. Of the 41 remaining in the sample, all but 1 returned with apparent eagerness for the second session, and the other required only very brief persuasion.

The adults, like the children, were not a random sample of the total population. Eight of them were completing or had completed the bachelor's degree requirements, and the other 13 had completed some graduate work (in medicine, optometry, social work, and clinical and experimental psychology).

No *S* showed any sign of physical or psychological disability. All were volunteers, and all were paid $5.00 in addition to whatever they received as incentives.

Apparatus

Physiological Recording Apparatus

For recording physiological tracings and RT data, three chart drives with ink-writing galvanometers were used. Chart Drive Number 1 was part of a custom-built Edin EEG-EMG which recorded the primary tracings of the EEGs and EMGs. High-pass filters designed to exclude frequencies below 20 cps were used to analyze EMG output (where the modal frequency is of the order of 40–50 cps). EEG was analyzed with the use of three band-pass filters, having fairly sharp cutoffs at their upper and lower limits (Ross & Davis, 1958). Band passes of 2–4 cps, 8–12 cps, and 17–28 cps were used, as were single frequency resonators for each of the frequencies 5, 6, and 7 cps. Electronic integrators, employing the condenser charge-discharge principle and connected in parallel fashion with the galvanometers of the EEG-EMG unit, summated the potentials over successive 4-sec. periods (Davis, 1956). The integrator discharges for EEG outputs and the integrated EMG potentials were recorded on Chart Drive Number 2. The ac outputs of all frequencies and bands of frequencies above 5 cps were also subjected to dc full-wave rectification, thus providing a separate, moment-to-moment record of amplitude of the activity at the various frequencies.

---

[3] The cooperation of A. C. Pitcairn is very gratefully acknowledged, as is that of the Protestant School Board of Greater Montreal.

On Chart Drive Number 3, in addition to EMG, previously mentioned, HR, respiration, and palmar conductance (PC) were recorded. HR was recorded by means of a continuous electrocardiographic tracing, and respiration was recorded with a Phipps and Bird pneumograph and recording tambour. PC was recorded by the monopolar method described by Malmo and Davis (1961).

Bipolar cellulose sponge electrodes, soaked in electrode jelly and attached to *S* by lastonet bands, were used for the EEG and EMG placements (Davis, 1959). The bilateral parietal EEG placement was located by marking a spot two thirds of the distance from nasion to inion along the center line of the skull and then placing the electrode centers 3 cm. to the right and left of this spot (for children the distance was 2.5 cm.). The EMG electrodes were placed to pick up activity of the left ("inactive") forearm extensor muscles, according to the instructions in the manual by Davis (1959). (The standard 2-in. distance between electrode centers was reduced in the case of children to 1.5 in.) The HR electrodes were also of cellulose sponge soaked in jelly and were held in place by elastoplast tape, on opposite sides of the chest, one just below the right collarbone, and the other just above the bottom of the left rib cage. The silver chloride electrodes used to record palmar resistance were placed on the palm and ventral forearm surface of the left arm and were held with lastonet bands. The pneumograph bellows were fastened about the chest wall.

RT Apparatus

The RT system was activated by a power switch, which started a cam rotating. The rotating cam activated microswitches which closed two relay circuits; one circuit started the simultaneous operation of a Standard Electric time clock (accurate to .01 sec.) and an electronic counter (Dymec Model Dy 2500 variable time base counter, accurate to 1 msec.); and the other circuit turned on the audio-oscillator, which presented the 1,000 cps stimulus tone.

The gain of the audio-oscillator was calibrated by *E* and another observer of normal hearing against the output of a Maico audiometer. The gain of the oscillator was set to deliver a stimulus tone 70 db. above *S*'s threshold, based upon an audiometric test.

Pressure on the reaction key broke the time circuit, stopping the clock and counter. The relay circuits connecting the reaction key with the timing devices had a release time of 4–8 msec. When the counter stopped, a Hewlett-Packard digital recorder (Model 560A) printed RT in milliseconds on a paper tape. Rotation of the cam broke the circuit presenting the stimulus tone after the tone had been on for .70 sec. Rotation of the cam continued to control the stimulus program and timing circuits until the power was turned off.

One rotation of the cam delivered 24 tones. The 24 intervals between the tones varied from 3.2 to 4.7 sec., with a mean interval of 4.0 sec. Half of the intervals was less than the mean, and half was greater. In the 24-interval sequence, long and short intervals occurred in random order.

The reaction key of this apparatus was attached to the right side of *S*'s chair, and its position was adjustable so that *S*'s index finger could rest comfortably upon it. One ink-writing galvanometer on each of the three chart drives recorded the onset and termination of each tone and each response.

Procedure

The following procedures apply to the children. Changes made in dealing with adults will be summarized below. The *S*s were run at 10 AM or between 1:30 and 2:30 PM. Every *S* came for two sessions, the second always occurring at the same time of day as the first and 1 week later, whenever possible. The 1-week interval actually obtained in only a little over half of the cases; other commitments of the parents or illnesses required the use of intervals as short as 4 days or as long as 6 weeks. The mean interval was about 10 days.

The physical setting was a two-room laboratory suite, the rooms being connected by a double door. During recording the doors were closed; but microphones, speakers, and a window allowed one-way visual and two-way auditory communication. The parent was comfortably seated in *S*'s room near her child; *E* and another operator manned three chart drives in the recording room where they had a clear view of the child. During the second session a female clinical psychologist was present in *S*'s room; her functions were to assist in putting on and removing electrodes, to make the child comfortable, and to assist *E* in giving instructions and in motivating the *S*s. The temperature of *S*'s room was maintained at 66° (±2°).

The first session, which lasted about 1 hr. 15 min., was designed to make the child familiar with the rooms, conditions, and personnel of the experiment so that physiological levels would not be affected by reaction to the novelty of the laboratory situation; to train him on the RT task until his performance became asymptotic; and to reward him sufficiently so that when he returned he would be a willing *S*. He was first given the audiometric test mentioned above. He was then made comfortable in a large semireclining chair, on the right arm of which was the reaction key, and gradually trained, over seven blocks of 15 reactions each, to work with a minimum of body movement and a maximum of speed.

Starting by pressing in response to *E*'s whistles, the child then pressed to a series of stimulus tones with his eyes closed (Block 1); and again, with the offer of a penny, if he were faster (Block 2); and then once more, this time with a blindfold and for two pennies, if he were faster than before (Block 3). Blindfold and earphones were removed; the child was offered candy; and *E* told him that he could play more games for bigger prizes, but that *E* had to put some wires on him so that he could tell what was happening to the child's heart beat and muscles and head while he played. The child was shown a drawer full of toys and asked to select one to try to win. In the drawer were story books, toy cars, watches, animals and dolls, marking pencils, cellophane tapes, nail clippers, paint and crayon sets and coloring books, bags of balloons, windmills, plastic pencil sharpeners, and a large variety of candies. (The most popular items were the toy watches, marking pencils, and cellophane tape.) The *E* showed the child the electrodes, said that their placement would not hurt, but added,

> Tell me if anything starts to hurt or bother you —then I'll be able to fix it.

Typically, the child, *E*, and his assistant engaged in small talk about school, teachers, and playmates during this period; and there were children's records, if the child wished to hear music. The child was then reseated, and the earphones and blindfold were replaced. Physiological recording commenced, beginning with a 1-min. resting period, followed by a "practice" block of reactions, three "game" blocks in which the child won 1¢, 3¢, and the toy, and a final resting period. Data from the first training session were not used in any analysis, except to establish the fact that RTs were asymptotic in both age groups.

The second session provided all of the data used in analysis. Its structure is illustrated schematically in Figure 1, and it is described in some detail below. The "periods" referred to in parenthesis refer to the periods of Figure 1.

1. Warm-up blocks (Periods A and B). Before the electrodes were attached, *S* completed two blocks of 18 reactions each "to get ready for the real games." These insured that he was warmed up, and provided a base line against which subsequent performance could be evaluated, especially as a check against any continued learning effects. Electrodes were then attached to him, and he was given five pennies as a reward for being so cooperative.

2. Initial rest (Period 1). After the electrodes were attached and recording started, *S* was asked to

> Just rest now. Sit very quietly and relax for just a minute. Pretend you're going to sleep.

At least 1 min. of record was taken.

3. Nonincentive (NI) Blocks 1 and 2 (Periods 2 and 3). *S* was told that these blocks, each of the standard 18 reactions, were for him to get used to the wires, and for *E* to get the machines fixed so games could be scored right. The standard instruction—"Get quiet and start to listen; get ready now; go."—was given in a monotone immediately following a reminder that it was not a game and that there was no prize. At the end of each block of reactions he was asked to sit quietly while *E* checked his machines. No score was given to *S*.

4. Incentive (I) Blocks 1 and 2 (Periods 4 and 5). The *S* was told that the machines were now fixed and ready for the first game, for a nickel. He was reminded of the rule that to win in any game, the score must be better than it had been in any previous game; he was told the score he must attain; and he was warned that he might lose some pennies if he were too slow. He was urged to get very comfortable, listen carefully, and press as fast as he could all the way through. He was then given the standard instructions in fairly dramatic intonation. At the end of the "nickel" game he was told to sit quietly while *E* added up the score and then was told that some of the presses were a little slow and that since his (fictitious) score was lower than he needed (the score announced was always too low, but dependent in part on the actual quality of the performance) two pennies would be taken from the five already given him. (The loss in this part of the session was introduced to convince *S* that he would not invariably win, since up to this point he had never failed.) The assistant then reassured *S* that he would very likely have a good game next time, when he played for two nickels, and urged him to listen very carefully and press very fast. The *E* reinforced this reassurance, gave the standard

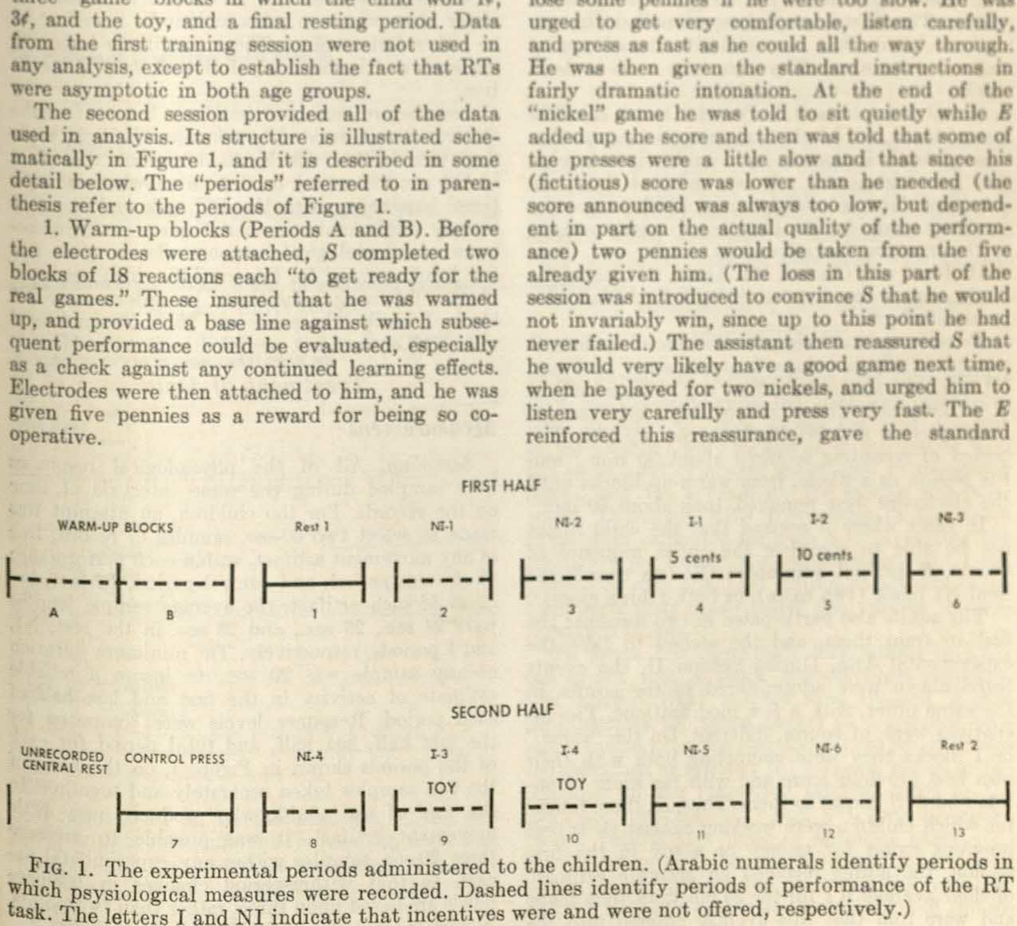


Fig. 1. The experimental periods administered to the children. (Arabic numerals identify periods in which psysiological measures were recorded. Dashed lines identify periods of performance of the RT task. The letters I and NI indicate that incentives were and were not offered, respectively.)

starting instructions, and announced to the child at the end of the second game block a "winning" score (fictitious, and based upon a scale of 0–100, with high scores "best"); the assistant then praised *S* and dropped two nickels into the money jar.

5. NI Block 3 (Period 6). The *E* interrupted the assistant to say that the machines needed fixing again, and a "practice" block was run as before.

6. Central rest period. The *E* suggested that *S* might like a rest and some candy. Blindfold and earphones were removed, and electrodes were checked, while *S* was shown his winnings and given candy. The *S* was then trained in pressing the key first at a pace called by *E* then at the same pace without *E*'s help. The pace approximated the average interstimulus interval of 4 sec.

7. Control pressing (CP; Period 7). The *S* repeatedly pressed the key for a minute, as he had been shown during the central rest period.

8. NI Block 4 (Period 8). This block was presented as necessary "machine fixing," as before.

9. I Blocks 3 and 4 (Periods 9 and 10). The prizes for these "games" were the toys which had been chosen by *S* at the end of the earlier training session. The *S* was, as before, reminded that he must do better than ever and given much urging and encouragement. The toys were won in both blocks.

10. NI Blocks 5 and 6 (Periods 11 and 12). The first of these two blocks was presented as another "fixing" trial. Following it, *E* said, with surprise, that time had almost run out, and that there was only time for one more practice block to get the machines ready for the little boy (or girl) coming in the morning (or afternoon).

11. Final rest (Period 13). Again, this lasted for 1 min., and it was terminated by a loud blast of a pneumatic horn.

To sum up the design for the second session, it was essentially a balanced presentation of I and NI blocks of reactions, so that any sessional effects of a reasonably linear nature (e.g., habituation) must affect I and NI blocks about equally. The period of recording required about 30 min.; and the session as a whole, from warm-up blocks until the electrodes were removed, took about 50 min.

In cases where it seemed that the child might not be able to complete the entire sequence of events, *E* did not administer CP (six cases), the final NI block (two cases), or both (three cases).

The adults also participated in two sessions: the first to train them, and the second to take the experimental data. During Session II, the events listed above were administered to the adults, in the same order, with a few modifications. The incentives were, of course, different. On the "game," or I blocks they were competing both with their own best previous score and with the score of another adult *S* known to them. Thus, on those blocks for which children were working against their best previous scores for money or a toy in the first session, the adults received an accurate estimate of their average RT for the 20 reactions they made and were told that this average entered into the average for the session as their competitive score. The score for the first I block was weighted by a factor of one in computing the sessional average, the last I block was weighted by a factor of three, and the intervening I blocks were weighted by a factor of two to parallel the gradual increase in incentive value offered the children. The prize for the first session winner was $1.00. For the second session, the loser of the competition in the first session was given a handicap of the number of milliseconds by which he had lost and was so informed, and the prize was $2.00. For the "practice" or NI trials, the adults were simply told to practice, without any explanation, and were given no knowledge of results.

The adults reacted 20 instead of 18 times per block simply because it was easier for *E* to find the average latency by dividing by 20. This procedure made each block 8 sec. longer for the adults. In addition there were two controls given to the adults which were not given to the children. The first was called control stimulus (CS) in which *S* just sat and listened to the sequence of 20 stimuli; and the second was called control response (CR) in which *S* responded to 20 stimuli as usual, but under instructions to relax and make no effort to be either fast or slow. These conditions lasted 80 sec. each and immediately followed the CP condition.

Just before summer holidays began for the children it was decided to ask some *S*s back, while they were still available, to determine the effect of CS condition and to determine whether additional experience with the RT task and the laboratory might make some difference in their patterns of physiological response. Seven girls and five boys returned for Session III, which consisted of the following conditions, in order: rest, NI block, I block (the prize was 25¢), CS, and final rest. EMGs were not recorded in Session III.

Treatment of Data

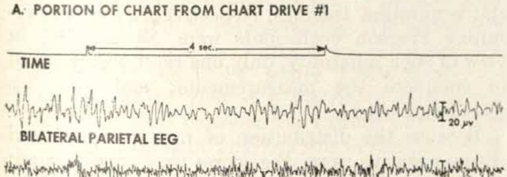
Measurement

Sampling. All of the physiological responses were sampled during the same intervals of time on the records. For the children, an attempt was made to select two 30-sec. samples of record, free of any movement artifact, within each performance block, rest period, and control period. In fact, because of such artifact, the average sample lengths were 24 sec., 26 sec., and 28 sec. in the rest, NI, and I periods, respectively. The minimum duration of any sample was 20 sec., to insure a reliable estimate of activity in the first and last half of each period. Response levels were computed for the first half, last half, and total period for each of the periods shown in Figure 1, on the basis of the two samples taken separately and together. In the case of the adults, who produced very little movement artifact, it was possible to measure three 20-sec. samples within any experimental period, so that within-period changes in response levels are based upon three, rather than two points. A portion of a record is illustrated in Figure 2.

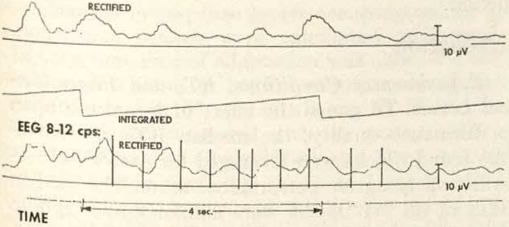
Autonomic Responses. HR was measured by counting the QRS deflections to the nearest quarter period within a sample, and expressing the result as a rate per minute. Respiration rate (RR) was measured by counting cycles to the nearest eighth of a cycle, expressing the result as a rate per minute. Palmar resistance was measured to the nearest 250 ohms. If the resistance tracing in the sample changed at a constant rate (as determined by a straightedge), one measurement was taken at the center of the sample. If the tracing was nonlinear, measurements were taken at 4-sec. intervals and averaged to establish the resistance for the sample. The resistance score of each sample within a period was then converted to a PC score (micromhos) by taking its reciprocal.

EMG. EMG was measured much like PC—i.e., if there was any apparent change in muscle tension within a sample, alternate 4-sec. integration deflections were measured by plastic scales which translated the deflection of the pen tracing from its base line directly into microvolts; if there was no apparent change in tension within a sample, only one central deflection was measured.

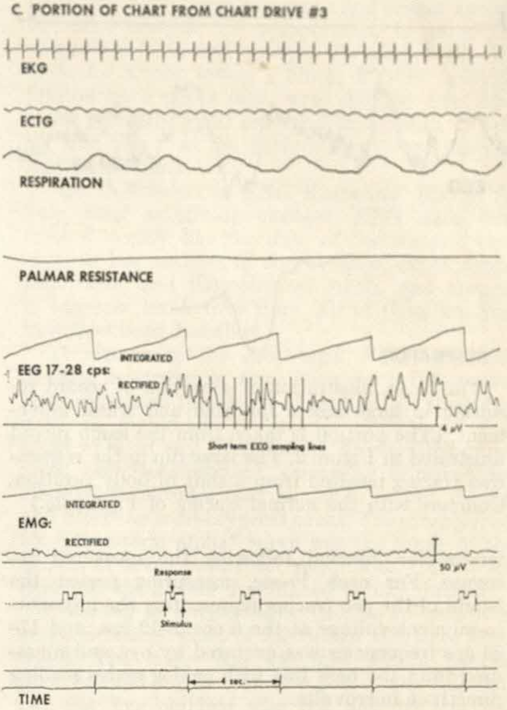
EEG. To assess the amplitude of the activity of the various EEG frequencies over time periods of 20 sec. or more, all of the integrator deflections (five to seven of them, depending upon the size of the sample) were measured in microvolts with the same kind of analog scales used in measuring EMG, and the average for the sample was computed. Dominant EEG frequencies in the alpha range were estimated for each *S* by averaging the frequency per second of approximately eight 1-sec. periods of such dominant sinusoidal activity taken from recordings made under the various motivational conditions.



2a. Shown are about 8 sec. of recorded primary EEG and EMG activity.



2b. Illustrated is the fractionation of the primary outputs into different frequencies, each represented both by integrator deflections occurring every 4 sec., and by a moment-to-moment rectified output.



2c. Shown are, from top to bottom, autonomic variables, beta and EMG outputs, and the S-R write-out. (The small vertical lines in 2b and 2c mark the seven time samples on the rectified EEG outputs which were used in the analysis of short-term amplitude changes associated with an S-R sequence.)

FIG. 2. Illustrations of the recordings used in measurement.

To assess *brief* amplitude fluctuations in EEG concurrent with a single tone-keypress sequence, the rectified EEG outputs were utilized. Figure 2b illustrates the seven phases of a rectified output measured for one reaction. These seven phases are described as follows: (*a*) a ½-sec. period at the midpoint of the foreperiod interval prior to the measured response; (*b*) and (*c*), two successive ½-sec. periods just prior to stimulus onset; (*d*) the latent, or RT, period; (*e*) and (*f*), two successive ½-sec. periods just following the response; and (*g*) a ½-sec. period at the midpoint of the in-

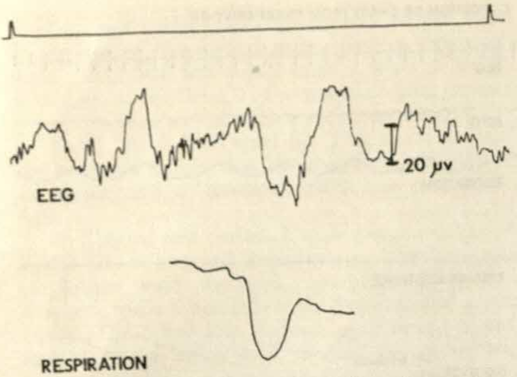


FIG. 3. An illustration of a portion of record resulting in large scores for head and trunk movement. (The portion is taken from the same record illustrated in Figure 2. The large dip in the respiration tracing resulted from a shift of body position. Compare with the normal tracing of Figure 2c.)

terstimulus interval following the measured response. For each ½-sec. measuring period the height of the pen tracing representing the moment-to-moment voltage at the 6 cps, 8–12 cps, and 17–28 cps frequencies was averaged by eye and measured from the base line with analog scales reading directly in microvolts.

Forty reactions in each record were selected for measurement, such that there were 10 reactions in each cell of a matrix, the dimensions of which were length of the interstimulus interval preceding the reaction (long or short) and speed of the reaction (fast or slow). On the average, of each set of 40 reactions measured, 24 occurred in NI periods and 16 in I periods. The records of only 21 of the children (10 girls and 11 boys) were measured. The remainder was not used solely for reasons of economy, since it required more than 24 hours of working time to process one record.

Performance. The average latency of reaction of the 18 reactions in a block was taken as the latency, in milliseconds, midway between the ninth and tenth scores in the order of latency and is referred to as the median RT of a block. Reactions of unusually short latency (shorter than half the latency of the median) were regarded as anticipatory reactions and arbitrarily assigned as reactions of *long* latency. Only 31 of these occurred in all the trials given the adults, and only 49 occurred in all the trials given the children.

Body Movement. Body movement is the most frequent source of artifact in the records; and it appears as a marked deviation in the pen tracings from their typical regularity, or as a departure of the actual from the calibrated base line of the tracing. Although there often occurred very brief, small body jerks which were especially correlated with the keypressing response (see the tiny "bumps" in the respiration tracing of Figure 2c), the kind of movement at issue is not response correlated, but rather a larger and longer lasting reorientation of the body or its parts. In observing the children, and in monitoring their records for such an artifact, there was a suggestion that the amount of what might be termed "restless movement" varied with the different experimental conditions; and it seemed desirable to assess the validity of this suggestion as well as the possible influence of such restless movement upon other responses.

To measure the movement, the pen tracings of EEG and respiration were utilized, since large head and body movements are easily seen in such tracings. For each *S*, for each tracing, and for each experimental period the number of ½-sec. periods in which a "base-line swing" or marked deflection occurred were counted (see Figure 3 for examples of such deflections). These frequency counts were adjusted for differences in the lengths of periods (e.g., the rest periods were typically shorter than the performance blocks). The small movements sometimes occurring in the S-R period were not counted.

Two raters practiced the counting system on two records and then rated 15 others independently. Scores of the two raters in three periods from each record (45 pairs of scores in all) were correlated for indications of "movement" in the EEG and respiration tracings, respectively; and the resulting Pearson coefficients were .83 and .89. In view of such reliability, only one rater was retained to complete the measurements; and only his scores were used.

Because the distribution of raw scores showed strong negative skew, the scores were transformed: Artifact in the EEG tracing was scored 1 or 0, indicating presence or absence of movement in the period; artifact in the respiration tracing was scored 2, 1, or 0, indicating a great deal of, a moderate amount of, or no movement in the period. There was no correlation within *S*s between the two scores of the two indicators of movement as they varied over experimental periods; yet, in the group, each score was highest in rest periods and lowest during I periods. Since this circumstance paralleled the paradigm of uncorrelated "tests" each correlating with a common "criterion," the two scores were combined by simple addition to form a single movement score for each *S* for each period.

Reduction

Experimental Conditions, RT, and Physiological Levels. To gauge the effect of incentives upon performance quality, the median RTs of each of the four I blocks were averaged for each *S* to construct an incentive performance score. The median RTs of all NI blocks were similarly averaged to form a nonincentive performance score.

Incentive and nonincentive scores were similarly calculated for the physiological and body movement scores and, in addition, average rest and control scores were computed.

To determine the relation between physiological levels and speed of reaction, the following procedure was applied to the data of each *S*, for each of his physiological measures, but will be illustrated only with reference to HR. The set of 10 HRs associated with the 10 performance blocks for any given *S* were rank ordered and then divided into three groups: the 3 highest rates were pooled as a "high-arousal" point, the 4 middle rates were considered together as an "intermediate-arousal" point, and the 3 lowest were combined as the "low-arousal" point. The performance score for each of these three arousal levels was the average RT of the 3, or 4, blocks whose associated HRs formed each arousal level.

Each *S* thus had three performance scores; and an average 3-point curve, based upon all *S*s in each group, was calculated separately at each age level. Each physiological measure served in turn as the index of arousal.

Concordance. The analysis of intraindividual physiological concordance was done by rank ordering, for each *S* and for each physiological measure, the levels of activity recorded during every experimental period of Session II, including rest and control periods, and then correlating the ranks of each measure with the ranks of all other measures. Thus, with reference to the 13 periods shown in Figure 1, the 13 corresponding HR scores were rank ordered and the ranks were correlated with the ranks of the 13 PC scores, EMG scores, etc., a process which yielded a matrix of intercorrelations for each *S* between the various physiological measures. The same analysis was performed on the adults' data, but there were 15 periods to be ranked. Finally, the 41 individual matrices of correlations of the group of children were averaged to form a matrix of average *intraindividual* correlations among measures, and a similar reduction was performed on the 21 adult matrices.

Adaptation. To assess changes in response levels occurring over the course of the entire session, an average of the physiological, RT, and movement responses was computed for each *S* for the five performance blocks in each half of the session. For example, the average HR of Periods 2–6 (see Figure 1) was calculated and compared with the average HR of Periods 8–12. A positive difference indicated a change downward in level, and a negative difference a change upward, over the session.

Changes in response levels occurring *within experimental periods* were calculated analogously, but the time unit of adaptation was about 1¼ min. rather than about ½ hr. For example, the average HR of the first 30 sec. of all six NI blocks was calculated and compared with the average HR of the last 30 sec. of all six NI blocks. The same calculations were made for the four I blocks. With adults, changes in response levels were represented by three, rather than two, points, it having been possible to sample three substantial portions of every block.

Block scores for body movement were too often zero to allow any subdivision into within-block parts. RT scores were numerous and precise enough to make possible the computation of three-part curves for both age groups. To obtain the RT score for a part within a block, RTs for a block, printed on a paper tape, were divided into three equal or nearly equal sets; and the median of each set was taken as the performance score for that part of the block.

Age Differences in EEG Responses. Except for very brief amplitude changes, EEG data were treated exactly like the data of the somatic variables in the analyses of concordance, experimental conditions, and physiological levels, and changes in response levels over time. All of these analyses have just been described.

To determine the short-term EEG amplitude changes occurring in association with RT, curves were calculated for each *S* at each of the three frequencies measured. The reader may recall that there were, in any given record, 40 reactions measured, each measurement consisting of seven sequential amplitudes read at points before, during, and after the tone-keypress event. For example, an *S*'s "short-term alpha" curve was the mean of the 40 sets of seven sequential amplitudes. Curves for beta and theta were determined in like manner.

Handling of Irregularities. Certain irregularities in the administration of the design will have been noted. The major one was that certain periods were omitted in the case of a few of the children. For the five children who did not perform the final NI block, the NI score is based upon only five instead of six NI periods. Any comparisons involving the CP period were made without reference to the data of the nine children to whom it was not administered. In calculating intraindividual concordances, there were three children for whom the rho's were based on only 11 ranks per measure instead of 13, and eight children for whom rho's were based upon 12 ranks per measure instead of 13. Appropriate reductions were made in the number of degrees of freedom used in evaluation of the significance of any average intraindividual rho of the group of children. In calculating direction of change of responses across the session in the case of the five children who did not perform the final NI block, the third NI block (Period 6) of the first half of the session was removed from analysis in order to render each half of the session comparable in terms of experimental conditions.

The adults had two control periods (CS and CR) not given to the children, and their session was thus a bit longer in terms of experimental periods. But this difference in length of session was for the most part cancelled because the children were given more time during the central rest period for eating candy and talking.

Other irregularities occurred, not in administration, but because *S*s in each age group responded differently. A majority of the adults tended to hold their breath when awaiting a stimulus tone; hence RR for adults could not be used as a valid physiological index. Such breath holding did not

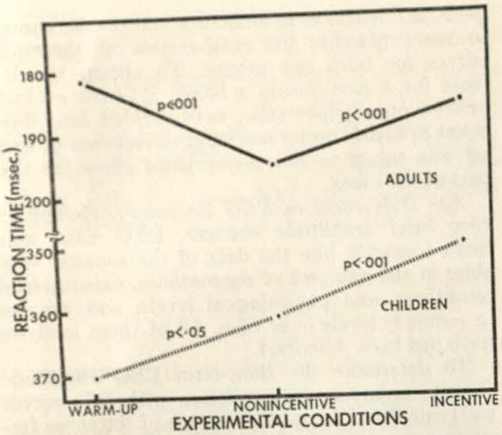


FIG. 4. Speed of reaction as a function of age and experimental conditions

occur in children. There was more artifact due to movement in the children's records than in the adult records. For this reason, as already noted, it was practicable to measure only two samples within experimental periods of the children's records, rather than three; and the physiological gradients within periods were thus based upon two points for children and three for adults. The responses to the horn blast terminating Session II involved an incidence of artifact in each group which was too high for the reliable measurement of anything but the nature of the artifact itself.

RESULTS AND DISCUSSION

The data are organized topically in the five sections which follow. The first section, and the core of the paper, deals with interrelations among experimental conditions, physiological measures, and reaction speed. The second section concerns concordance; the third deals with adaptation; and the fourth with the effects of interstimulus (preparatory) interval on RT. A discussion follows the presentation of results on each topic. Because results in one section bear to some extent upon results in another, an attempt has been made to discuss one set of results in terms of another whenever such cross-referencing seemed appropriate. The discussion of EEG data from the first three sections has been for the most part put off until the fifth and final section.

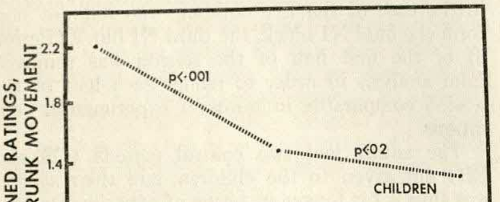


FIG. 5. "Restless" movement of body parts as a function of age and experimental conditions.

The following distinctions have been used in referring to the various measures: the term "physiological" refers to all of the physiological measures; the terms "somatic" or "bodily" refer to all physiological measures except EEG; the term "autonomic" refers to HR, PC, and RR; "performance" refers to responses made in the execution of a task; "motor" or "overt behavioral" refers to all skeletal motor responses, whether or not made in the execution of a task.

Experimental Conditions, RT, and Physiological Levels

Effects of Experimental Conditions upon Overt Responses

RT. The average RTs of the adults and of the children under each of the conditions requiring serial RTs (two preliminary warm-up blocks, six NI blocks, and four I blocks) are illustrated in Figure 4. The adults were at their fastest, and the children at their slowest in the warm-up conditions. This result accounted for most of the significant Age × Conditions interaction when the RT scores were subjected to a Type I analysis of variance ($F = 5.82$, $df = 2/116$, $p < .001$; Lindquist, 1953). There was a marked improvement in speed of reaction in each group ($p < .001$ for each group) when incentives were offered; and there was no significant difference between the groups in the reliability of this improvement, which occurred in 18 of the adults and 27 of the children. A comparison of the change scores, NI minus I, of the groups yielded a t of .68, not significant.

Body Movement. Changes in the amount of body movement as a function of experimental conditions are illustrated in Figure 5. It is clear that the children made a great many more noninstrumental movements than the adults did in every condition, a result in keeping with the informal observation that the young child does not "keep still" very much except in states of sleep or of strong attention. Body movement declined in incidence progressively for both groups as conditions changed from rest to NI to I, and the decline was significantly steeper for children than for adults. This interaction, however, may be ascribed to age differences in the intraindividual range of movement scores (see Table 6); the proportions of *S*s showing the illustrated change are not significantly different from one age group to the other: 14 of the 21 adults and 33 of the 41 children showed less movement during NI blocks than during rests; 12 of the adults and 28 of the children showed less movement during I blocks than during NI blocks.

Effects of Experimental Conditions upon Physiological Levels

The results of the effect of the rest, NI, and I conditions upon the levels of different physiological variables are summarized in Tables 1, 2, and 3. Table 1 presents the mean level of each variable in each condition, for the ages separately. Table 2 presents a summary of the analyses of variance which were applied to the data of Table 1. These analyses were of two kinds: *S* × Treatments designs were employed to analyze each age group separately, and Type I designs (Lindquist, 1953) were used to analyze the data of the combined groups. Table 3 shows for each age group separately the amount and the significance of the change in the level of each physiological measure from the rest to the NI conditions, and from the NI to the I conditions. The table also indicates, in the columns labeled Interaction, whether the adult group demonstrated a reliably greater change in levels from one condition to another than the children did. (Children in no case changed in level from one condition to another significantly more than the adults.)

Changes from the Rest to the NI Conditions. Table 3 indicates that the PC and EMG scores of children and adults rose significantly and in a parallel way (i.e., no significant interactions) from rest to NI. In the cases of the HR and EEG 8–12 measures, there were significant differences between the ages in the recorded changes from rest to NI. For EEG 8–12, this interaction is due solely to an age difference in the *direction* of change (a phenomenon which will be discussed later) and not to any difference in *magnitude* of change; there was no age difference in the absolute change in EEG. For HR, the interaction seems to have resulted from the fact that the range through which HR varies within the adult is much larger than the range in children (see Table 6), a situation which permits greater change to occur in adults. When the age difference in variability was eliminated by recourse to nonparametric analysis, there was no reliable difference between the proportions of

TABLE 1

Mean Values of all Physiological Measures in the Rest, NI, and I Conditions for Adults and Children

Physiological measures	Age group					
	Adults			Children		
	Rest	NI	I	Rest	NI	I
HR (bpm)	72.86	77.43	81.02	92.59	94.02	95.03
PC (micromhos)	95.83	101.61	106.42	172.86	188.98	191.13
RR (cpm)	15.38	—[a]	—[a]	19.13	21.06	21.32
EMG (microvolts)	5.35	7.94	10.58	5.29	7.32	7.21
EEG 8–12 (microvolts)	7.86	7.15	6.81	9.86	10.44	10.52

[a] Adults so often held their breath during performance that RR score was not used.

TABLE 2

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE OF *Ss'* SCORES FOR REST, NI, AND I CONDITIONS

Source	*df*	*F*
HR		
Conditions		
Adults	2/40	24.16***
Children	2/80	24.78***
All *Ss*	2/120	47.31***
Age	1/60	36.49***
Age × Conditions	2/120	13.89***
PC		
Conditions		
Adults	2/40	13.54***
Children	2/80	17.50***
All *Ss*	2/120	23.55***
Age	1/60	18.16***
Age × Conditions	2/120	.71
RR[a]		
Condition		
Children	2/80	63.18***
Muscle Tension		
Condition		
Adults	2/40	131.29***
Children	2/80	10.52***
All *Ss*	2/120	153.34***
Age	1/60	.08
Age × Conditions	2/120	43.75***
EEG 8–12		
Conditions		
Adults	2/40	6.28**
Children	2/80	6.25**
All *Ss*	2/120	.64
Age	1/60	8.25**
Age × Conditions	2/120	12.12***

[a] See Footnote a, Table 1.
** $p < .01$.
*** $p < .001$.

adults (18 of 21) and of children (30 of 41) whose HRs rose. The RR of the children also rose very significantly.

It appears that when allowances are made for age differences in range of variation of HR and in direction of change of EEG 8–12, the children and the adults changed in a similar way in physiological levels as conditions changed from rest to NI. It is at least possible to say that physiological measures obtained from children can discriminate between certain conditions as effectively as they do in adults.

Changes from the NI to the I Conditions. The data on the physiological changes from the NI to the I conditions are also presented in Table 3. Among adults, there were significant changes in the direction of greater excitation in every measure; but among children, only HR and RR differentiated the two conditions to a significant degree. On the four variables for which age comparisons could be made (i.e., excluding RR) either there was a greater change in adults from NI to I than there was in children, or there was a significant change in adults where none existed in children. This summary required modification, however, in the case of HR: the significantly greater change in adults may again have been a function of greater within-adult range, because the analogous nonparametric test for interaction revealed no difference in the proportion of adults (17 of 21) and of children (31 of 41) whose HR rose with incentives.

Figure 6, based upon EMG data, can be viewed as a summary illustration of the effects of experimental conditions upon physiological levels. It includes the two modal effects: first, the similar degree of change in the two age groups from rest to NI; and second, the divergence of the two age groups in their degree of change from NI to I.

It was shown in Tables 1 and 2 that the children had higher overall levels of autonomic activity than the adults. The possibility existed that those children with adultlike low levels of activity might appear more similar to adults than to other children in showing much stronger elevations of such activity under the I condition than under the NI condition. The 10 children with the lowest resting levels of HR (average 80 beats per minute or bpm) and the 10 with the lowest resting levels of PC (average 90 micromhos) were separately analyzed. The mean change in HR from NI to I was .78 bpm, less than the change for the group of children as a whole, and not significant. The mean change in PC from NI to I was 4.5 micromhos, also less than the corresponding change for the entire sample of children, and not significant. Thus, even those children who appear similar to adults in resting level are not similar to them in the patterns of change of HR and PC from the NI to the I conditions.

TABLE 3

Effects of Conditions upon Physiological Measures: Mean Differences and Their Associated Significance Levels between Pairs of Conditions for Each Age

Physiological measures	Conditions compared					
	NI minus rest			I minus NI		
	Adults' $\overline{D}$	Children's $\overline{D}$	Interaction p	Adults' $\overline{D}$	Children's $\overline{D}$	Interaction p
HR (bpm)	4.57^{****}	1.43^{****}	$<.05$	3.59^{***}	1.01^{****}	$<.01$
PC (micromhos)	5.78^{*}	16.12^{****}	*ns*	4.81^{***}	2.15	*ns*
RR (cpm)	—[a]	1.93^{****}	—[a]	—[a]	$.24^{*}$	—[a]
EMG (microvolts)	2.59^{***}	2.03^{****}	*ns*	2.64^{**}	−.11	$<.01$
EEG 8–12 (microvolts)	$-.71^{*}$	$.58^{*}$	$<.01$[b]	$-.34^{***}$	.08	*ns*

[a] See Footnote a, Table 1.

[b] The interaction is due to the difference between age groups in direction, not amount of change.

$^{*}p < .05$.

$^{**}p < .02$.

$^{***}p < .01$.

$^{****}p < .001$.

Comparison of the Rest and Control Conditions. The control condition in the children was designed to assess the effect of keypressing alone upon levels of the physiological measures. When the CP levels were compared with rest levels, there were no significant, or even suggestive, differences for HR, PC, and EEG 8–12; there were, however, significant differences in RR ($p <$ .05), and EMG ($p < .01$), the control levels being higher. The adults participated in three different control periods. The three were very similar to each other in levels of the different measures, and there was no difference between the control and rest conditions for any measure except, as with the children, RR ($p < .01$).

There was a very obvious difference between the age groups in the performance of the keypressing task in the CP condition. All *S*s were instructed alike to press at a slow, regular pace approximating the mean interstimulus interval of 4 sec. On the average the children made about 62 irregularly spaced presses in a minute, while the adults made about 14 very regular ones. In view of the already established failure of the EMGs of children to differentiate the I from the NI conditions, even in direction of trend, and in view of the fact that EMG during the frequent responding of the CP conditions was at the second highest level of any of the 13 periods measured in the case of children, the possibility seems strong that the left arm extensor EMGs in the children were a function primarily of the right ("active") arm flexor activity.

Results from Session III

The data from Session III were analyzed like the data of Session II. A comparison of the two sets of analyses is presented in Table 4. It is important to note that the values of the statistic *T* (Wilcoxon, 1949) listed in the table for Session II are based *only* upon the data of the 12 children who returned for Session III. The trends which appeared from the data of Session II also appear again in Session III, i.e., there were still significant differences between rest and

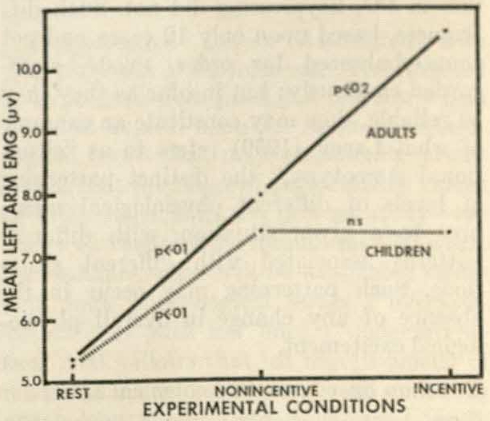


Fig. 6. Muscle tension (EMG) as a function of age and experimental conditions.

TABLE 4

WILCOXON *T*s BETWEEN PAIRS OF CONDITIONS IN SESSIONS II AND III, FOR SCORES ON PHYSIOLOGICAL MEASURES, MOVEMENT, AND RT

Measure	Conditions compared					
	Control minus rest		NI minus rest		I minus NI	
	II	III	II	III	II	III
HR	39.5	48.0	0***	20.0	16.0	19.0
PC	23.0	10.0**	11.0*	5.0***	39.0	17.0
RR	0***	28.0	0***	6.0***	24.0	54.0
EEG 8–12	39.0	40.0	27.0	9.0**	26.0	53.0
Movement	7.0^{a**}	0^{a***}	0^{a***}	2.0^{b***}	12.5^{b}	—^{c}
RT	—	—	—	—	12.0*	6.0***

Note.—*T*s represent the sum of negative ranks when physiological measures are the basis of comparison. *T*s represent the sum of positive ranks when movement and RT scores are the basis of comparison.

$N = 12$ in all comparisons, except as noted.

[a] Because of a zero differrence, $N = 11$ for the test.

[b] Because of two zero differences, $N = 10$ for the test.

[c] Because of 8 zero differences, no test was possible.

* $p < .05$ level, 2-tailed test.

** $p < .02$ level, 2-tailed test.

*** $p < .01$ level, 2-tailed test.

NI in the physiological and movement variables, there were still no differences between the I and NI conditions, and the presence of the incentive resulted in much shorter RTs. Thus, the finding that the introduction of incentives improved performance markedly while having only slight effects upon physiological levels was a reproducible one.

The control condition in the third session was not keypressing but simply listening to the stimulus tone. Neither control condition differed from the resting condition when HR and EEG 8–12 were the variables of comparison. But where keypressing resulted in elevations of RR, listening did not; and where listening resulted in elevation in PC, keypressing did not. Such differences, based upon only 12 cases and not counterbalanced for order, must be regarded cautiously; but insofar as they may be reliable, they may constitute an example of what Lacey (1959) refers to as "situational stereotypy," the distinct patterning of levels of different physiological measures to a given situation, with different patterns associated with different situations. Such patterning may occur in the absence of any change in overall physiological excitement.

Reaction Speed and Physiological Level

The results concerning the relation between physiological measures and RT are based only upon those 10 periods during which *S*s reacted to blocks of stimuli in the RT task. Up to this point it has been shown that the RT performance of adults and children improved in similar ways when incentives were offered. In adults, the increase in speed was accompanied by a parallel increase in the levels of measured physiological activity and a decrease in EEG amplitudes; but among the 41 children the only physiological changes accompanying the improvement in performance were increases in HR and RR.

Reaction Speed at Low, Middle, and High Physiological Levels. The question now is, what is the nature of the function when RT is plotted directly against varying levels of the different physiological functions? Figures 7 and 8 illustrate such functions, based upon levels of HR and PC within each age group. Note that the three points on the abscissa are based upon different levels of physiological activity in the same *S*s; they are *not* based upon different groups of *S*s. The difference between the high and low levels of HR averaged 8.5 bpm in adults, and 6 bpm in children, or about 75–80% of the total intraindividual range in each group. The difference between the high and low levels of PC averaged 32.5 micromhos in adults, and 62 micromhos in children, or about 93% of the total range in each group (see Table 6).

HR was chosen for illustration because

it increased, as did reaction speed, very significantly in both groups under I conditions. Conductance was chosen because its average intraindividual range in children was relatively large, so that there was "room" within its range of variation for a systematic relation with performance to occur, if any existed. It is clear, however, that the variation in performance did not relate in any way to the three levels of HR or of PC within the average child. Within the average adult, on the contrary, there were significant monotonic increases in reaction speed as a function of increasing levels of HR and PC. The relations between levels of all other physiological variables and RT were not significant in either age group.

In order to clarify the analyses next to be presented, a brief discussion of the results plotted in Figures 7 and 8 is offered here. It is possible that the relations between HR and reaction speed or between PC and reaction speed which existed among adults were artifactually generated because the variance of each set of measures (autonomic and performance) may have been chiefly determined by variation in incentives. Furthermore, it is also possible that the apparent *lack* of any relation between physiological levels and performance efficiency among children was a sampling artifact. That is, the experiment may have sampled so limited a portion of the possible range of physiological levels of each child that the flat curves of Figures 7 and 8 are actually composites of a number of individually lawful monotonic relations between activation and performance, each one of which represents only a part of the postulated nonmonotonic, inverted **U** relation. The activation levels of some children may have been entirely below the "optimum" level for the task; the activation levels of other children may have been entirely above such an optimum level; and the result of summing over *S*s would be the illustrated flat curves. The analyses which follow were done to answer unequivocally the question whether levels of physiological activity were, in fact, related to reaction speed.

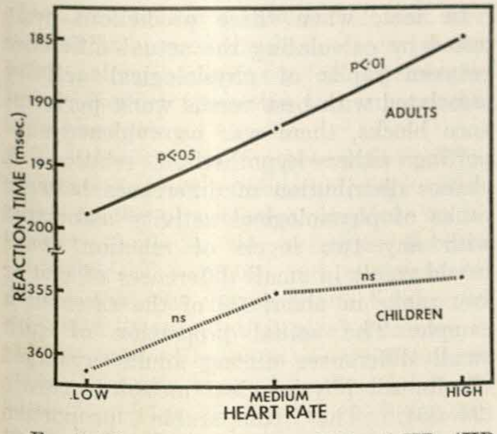


Fig. 7. The relation between RT and HR. (HR has been divided into three levels—low, medium, and high—per *S*.)

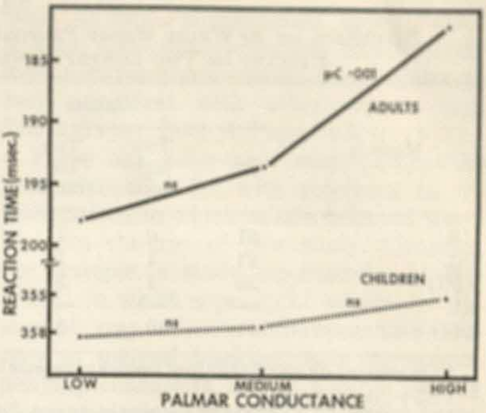


Fig. 8. The relation between RT and PC. (PC has been divided into three levels—low, medium, and high—per *S*.)

Physiological Levels Associated with Worst Blocks of Performance. The inverted **U** hypothesis does predict a relation which does not depend upon the range of physiological levels that may empirically occur. Given that the range of performance values is wide (as it is in the present study—see Table 6) then, by the hypothesis, the worst performance value will be associated *either* with high physiological levels ("overactivation") *or* with low ones ("underactivation"). It follows that, of the 10 blocks of reactions made by each *S*, the block in which the slowest reactions were made should have been associated more often

TABLE 5

FREQUENCY OF *Ss* WHOSE WORST PERFORMANCE BLOCK OCCURRED AT EITHER THE TWO HIGHEST OR TWO LOWEST RANKS OF EACH PHYSIOLOGICAL MEASURE

Measure	Adults				Children			
	N[a]	Two high ranks	Two low ranks	Total extremes	*N*[a]	Two high ranks	Two low ranks	Total extremes
HR	21	0	7	7	40	10	10	20
PC	21	1	7	8	40	8	6	14
EMG	20	1	8	9	40	9	5	14
RR	—[b]	—[b]	—[b]	—[b]	39	9	10	19
EEG 8–12[c]	21	0	3	3	40	9	6	15

[a] The observed *N*s vary slightly from measure to measure because of the occasional failure in the recording in one physiological channel.

[b] See Footnote a, Table 1.

[c] The highest rank for EEG 8–12 was assigned the period of lowest amplitude.

than chance expectation with *either* the highest *or* the lowest levels of a physiological indicant. The frequency with which the worst performance block was accompanied by 1 of the 2 lowest (Ranks 9 and 10) or 2 highest (Ranks 1 and 2) ranks of every physiological measure was tabulated, and the results appear in Table 5. The expected frequency, on a chance basis, of the worst performance being associated with 4 of 10 physiological levels is about 16 in the group of children. The highest observed frequency tabulated for the ranks 1, 2, 9, and 10 of any physiological measure was 20, a nonsignificant departure from expectation. Among adults, the expected frequency was about 8.4, and the highest observed frequency was 9, again not significant. Worst performance blocks occurred among children about as often at the 2 highest levels of a physiological response as at the 2 lowest levels, all measures considered; in adults, the worst performance block occurred far more often at the low than at the high levels of a physiological measure, all measures considered. It thus appears that the worst score of 10 highly variable performance scores may and often does occur at fairly moderate levels of recorded physiological activity.

Discrimination of Extremes of Performance Quality by Levels of Physiological Activity. If one hypothesizes a relation between level of physiological activity and quality of performance, then regardless of whether the hypothesized relation is monotonic or nonmonotonic in form, one must assume that the levels of physiological activity associated with the best and worst blocks of performance ought to be fairly widely separated along the dimension of variation of a recorded physiological activity. Specifically with regard to the present design, which recorded 10 ranks of activity (one for each performance block) for each physiological measure, it follows that the rank assigned to a measure of physiological activity associated with the best performance block should differ from the corresponding rank assigned to the worst performance block. This difference should be of about 4 or 5 ranks, if the hypothesized relation is of the inverted U form, or about 8 or 9 ranks, if the hypothesized relation is monotonic.

In fact, when these predictions were tested by calculating the actual difference between ranks of physiological activity associated with best versus worst performance blocks, there was no evidence supporting either hypothesized relation. A chance distribution of differences between ranks of physiological activity associated with any two levels of reaction speed would result in small differences of one or two ranks in about .38 of the cases in a sample. The actual proportion of such small differences among adults averaged .29 for all physiological measures (range .24–.35). The comparable proportion among children was an average of .31 (range .23–.38). This incidence of very small differences of only one or two ranks between the physiological levels associated with opposite extremes of performance efficiency was slightly, although not significantly, less than chance expectation in each group. There is, thus, no support for the expectation that fairly large differences in physiological activity will be associated with extreme differences in efficiency of performance.

Discussion

The results of this section may be itemized as follows. First, reaction speed was strongly associated with the major independent variable, presence or absence of incentives, to about the same degree in both adults and children. Second, in adults there were strong relations between incentive levels and every physiological measure, but in children the relation existed only for the two measures, HR and RR. Third, in adults, there were significant relations (monotonic increasing) between reaction speed and the levels of both HR and PC; but among children, there was no relation between reaction speed and any physiological measure. Fourth, from theoretical assumptions which specify monotonic or inverted U relations between level of physiological activity and efficiency of performance, two predictions were drawn: One specified that the block of slowest average reaction speed should be associated with either extreme of the activity of a physiological measure; the other specified that the blocks of fastest and slowest average reaction speed should not occur at near points along a dimension of physiological levels. Neither prediction was supported in either age group.

The chief question in these data is: What accounts for the considerable age differences in the physiological response to incentives and in the relations between physiological response and performance? Recall that the intraindividual range of variation of the physiological measures in children was adequate to permit systematic changes of the kind shown by adults to occur, and that, in fact, the children did show physiological changes as great as those of adults when conditions changed from rest to NI. The children were not, then, at any "physiological upper limit," though they were operating at relatively higher physiological levels than adults (a fact consistent with other reports—e.g., Lucas, Pryor, Bost, & Pope, 1933).

These age differences may perhaps be best accounted for with reference to investigations in which adults behaved much like the children of this study. Consider, for example, a study reported by Malmo (1963) in which a group of *adults* demonstrated very reliable differences in accuracy of manual tracking, with no accompanying changes in any of a large battery of physiological measures. These results, which are similar to those found in the child *S*s of the present investigation, were attributed to the differences in instructional set given to Malmo's *S*s: In one condition ("unified set") they were told to track manually to an auditory cue; in the other condition ("divided set") they were also told to track manually to the auditory cue, but to expect at any moment to have to track, in addition, pedally to a vibratory cue. It was, of course, the divided-set condition in which the poorer performance occurred, even though, for the experimental periods compared, *S*s were never actually required to track pedally. An interpretation of the decremental effects of divided set is that such a condition prevents *S* from fully organizing a dominant sequence of responses because he must maintain another set of responses in partial readiness—there is conflict between competing organizations.

Something akin to such a state of conflict seems to have been present among the children during the NI condition. It is well known that children are not as effectively motivated by delayed reward as adults are (Bialer, 1961). (An example in the present study is the performance of the two groups in the preliminary warm-up blocks, for which no incentives were offered: the adults performed at their best, the children at their worst.) Thus, though the adult may have interpreted the NI blocks (presented as "practice" blocks) as integral parts of a goal-directed sequence, the children very likely did not. Rather, it seems more probable that the NI condition was for them a state of conflict between a tendency, on the one hand, to obey *E*'s instruction to sit quietly and work, and a tendency, on the other, to escape from an unrewarding and physically restricting task. The evidence for these inferences in the behavior of the children is the frequent incidence of restless movement during NI trials, interpreted here as a symptom of divided attention or of inattention to the task in much the same sense as Luria (1932) interpreted irregular motion as evidence of conflict.

Malmo's study represents a case in which conflict or divided set was induced in normal adult *S*s. There are other reports of results obtained in adult samples which parallel those found here among children. Malmo and Shagass (1949) found that among anxiety neurotics and early schizophrenics the responses chiefly differentiating periods of stress due to pain stimulation from periods without stimulation were not autonomic and EEG responses, but rather movement responses of finger and head, neck muscle potentials, and respiratory irregularities of the sort used in this study as an index of trunk movement. In a normal control group, however, no such "splitting" of overt and covert responses occurred; diffuse motor responses were inhibited, and motor control was maintained. A recent study by Mirsky and Cardon (1962) tested eight normal adults on vigilance tasks under the effects of chlorpromazine and under the effects of sleep deprivation. Although both manipulations resulted in performance decrement, the decrement under sleep loss was accompanied by systematic changes in physiological levels, but the decrement under chlorpromazine was not. These studies are instances, again, in which clinically or experimentally disturbed adults, like the sample under divided set in Malmo's study or like the children in this one, have shown changes in perceptual-motor or other overt behavioral responses which had no accompaniment in physiological changes. The source of the conflict or divided set differs from study to study; but it is contended here that youth, psychosis, some drugs, and conflict-producing instructions have the common effect of rendering the organism highly susceptible to task-irrelevant cues and that such an effect is capable of making changes in overt responses clearly independent of changes in physiological responses.

On the other hand, physiological and motor variables are sometimes strongly related (Duffy, 1962) as in the case of the covariation of HR and reaction speed among adults of this investigation. Is it then possible to conclude that under normal conditions there is some direct relation between activation and performance? The view here is that such a conclusion is neither necessary nor justified because, in the present case, it was the manipulation of incentive level, not of HR, which was the independent variable.

The physiological variables in this study, as in others similar to it (Belanger & Feldman, 1962; Freeman, 1940; Stennett, 1957b) are measured concurrently with the performance measure; they have no priority in time. This being the case, one is not justified in assuming that variation in level of physiological activity is producing variation in performance any more than one is justified in assuming the converse.

In view of the considerations just stated, perhaps the simplest way to conceptualize the results of this study, or any study in which covert physiological and overt motor variables are concurrently measured, is to assume that both the overt and covert behaviors are affected chiefly not by each other, but by whatever antecedent manipulations have been made. The argument is not that physiological and perceptual-motor responses are completely independent; they both occur, after all, within a system in which interactions of all kinds are characteristic, and it is obviously true that insults to the head and body can directly produce decrement in performance. What is being maintained is that under normal conditions the two sets of measures can be parsimoniously considered as responses which are largely independent, not related by any given function, and not so organized that one set has any special influence over the other. A well-known view

(Bindra, 1959; Duffy, 1962; Hebb, 1955; Malmo, 1959) is that the physiological measures may represent, with varying degrees of validity, the hypothetical construct known as activation, and that activation operates upon efficiency of performance. But it is most unlikely that this view can be tested with the peripheral measures usually used as indicators of activation because the investigator can never be assured that his *S*'s response to stimulus change is not itself producing independent, though apparently correlated, changes in both the perceptual-motor behavior and the physiological indicants of activation.

The interpretation of the data offered here as an alternative to an explanation based upon activation is virtually identical to the "multiple-concurrent response" model offered by Eriksen (1956) as an interpretation of the "subception effect" (Lazarus & McCleary, 1951) to replace the interpretation based upon the notion of "unconscious discrimination." Similar interpretations, discussed with reference to the relations between physiological activity and quality of performance, have been put forward recently by Church (1962) and Mirsky and Cardon (1962).

In the discussion above, the differences between children and adults have been interpreted, not as maturational differences of a gross physiological sort, but as a difference in psychological response to the situation, in which the children were especially prone to develop conflict. But there was one result in which maturational differences seem to have been directly responsible for the group differences. It was the differential response of EMG. In children, muscle tension was actually less during I than during NI conditions; yet their EMG scores during the control, simple keypressing condition were very high, at the same time in which autonomic measures indicated relaxation. These results are difficult to explain without reference to the fact that the muscles of 6-year-olds are very different from those of adults, both in terms of composition and of proportion of body weight (Zubek & Solberg, 1954). Central control of muscle action seems also to be less well differentiated in the young child than in the adult. Thus, a recent developmental study by Missiuro (1963) reports that flexion of the right biceps is accompanied, in the young child but not the adult, by strong EMG activity in the left extensor (triceps). Such an effect helps explain why there was so much measured left-extensor EMG during the otherwise relaxing CP condition: The children were making a great many flexor movements with the right hand.

A final result has some relevance to the choice of autonomic measurement in research. HR was the most reliable physiological differentiator of experimental conditions in each age group. This is a result which has been found previously in this laboratory in studies of adults (MacNeilage, 1961; Schnore, 1959; Stennett, 1957b;[4] Weiffenbach, 1962) and which has also been recently reported by Lacey, Kagan, Lacey, and Moss (1962). Nevertheless, PC has been the classical measure of arousal; and both Haywood and Hunt (1963) and Mandler, Mandler, Kremen, and Sholiton (1961) have in recent papers argued the superiority of PC over HR as a measure of arousal. It is probably true that PC is more sensitive than HR to environmental change—PC is a well-known part of the orienting response, and it will rise briefly even if the environmental change is an instruction to relax. But this indiscriminate sensitivity to change renders PC somewhat less effective than HR in indexing the physiological effects of motivational conditions which are maintained over appreciable periods of time. Many studies of arousal do not, after all, concern themselves especially with the orienting reflex.

To summarize this section, with reference to the general question of the age differences discovered, it is the writer's view that, with the possible exception of the EMG responses, the differences are

---

[4] The reference is to HR data which Stennett gathered for his dissertation, but did not analyze. The author analyzed them in the course of this investigation.

TABLE 6

COMPARISON OF ADULTS AND CHILDREN WITH RESPECT TO AVERAGE INTRAINDIVIDUAL RANGE OF VARIATION OF MEASURES

Variable	Range					
	All experimental periods			Performance blocks only		
	Adults	Children	D	Adults	Children	D
HR (bpm)	14.4	8.8	5.6**	11.9	7.6	4.3*
PC (micromhos)	41.8	78.0	36.2***	34.7	66.9	32.2***
RR[a] (cpm)	—	5.2	—	—	3.9	—
EMG: Left-Forearm Extensor (microvolts)	15.2	13.5	1.7	14.0	12.2	1.8
EEG 8–12 (microvolts)	6.5	8.0	1.5	4.6	6.4	1.8**
Movement (Rating 0–3)	1.7	2.4	.7*	1.2	2.1	.9***
RT (msec.)	—	—	—	46.9	107.6	60.7***

[a] See Footnote a, Table 1.
* $p < .05$, 2-tailed test.
** $p < .02$, 2-tailed test.
*** $p < .001$, 2-tailed test.

evidence, not so much of age differences in physiological function, as of differences in the way in which the child and the adult responded to the experimental conditions. The adult differed from the child with respect to characteristics such as tolerance for delay of rewards, capacity to inhibit competing responses, capacity to remain still, desire to impress *E*, strength of the motivation to achieve, and so on. These characteristics fall, to use a classical distinction, more under the heading of learning than of maturation.

Concordance of Physiological Measures

Before comparing the within-*S* covariation of physiological response measures in children with the corresponding covariation for adults, we must look at the average range of variability of the measures within *S*s over all the experimental periods from the initial to the final rest periods, in order to determine whether the two age groups were sufficiently similar in this regard to permit valid comparisons of the correlations between physiological measures. If, for example, PC and HR both varied to a larger degree from period to period within *S*s in the adult group than they did in the case of the children, a larger correlation between those physiological measures in adults might be partly accounted for by the fact that the size of observed correlations depends upon the ranges of the variables measured. Table 6 presents the mean intraindividual range of each measure for each age group. The differences between the ages with reference to the intraindividual range of variation of a measure was determined for all measures by the Mann-Whitney *U* test or, in the case of movement, by a median test. The major age differences in the intraindividual range of variation occurred in HR, which varied more widely within adults than within children, and in PC, which varied more widely within children than within adults.

Table 7 presents the average intraindividual rho's between pairs of the physiological variables for each age group. The significance levels attached to the average rho's shown were determined by transforming every coefficient to a *z* score (i.e., Fisher's transformation), taking the mean of the *z* scores and determining the significance of that mean. Age differences between the sizes of corresponding average coefficients were evaluated by median tests, the dichotomies being adults versus children, and above- versus below-median coefficients. In each group there was a well-marked cluster of highly intercorrelated EEG amplitude scores, and the groups did not differ in the size of these correlations. There was also a cluster in

TABLE 7

AVERAGE INTRAINDIVIDUAL CORRELATIONS (RHO'S) BETWEEN PAIRS OF PHYSIOLOGICAL MEASURES

Measure	HR	PC	RR[a]	EMG	EEG 8-12	EEG 17-26	EEG 6
HR		.46***[b]	—	.45***[c]	−.16**	−.11*	−.11
PC	.12*		—	.35***[d]	−.16**	−.08	−.02
RR	.36***	.17**		—	—	—	—
EMG	.08	.07	.19**		−.05	−.07	−.02
EEG 8-12	.00	.04	.03	.00		.55***	.66***
EEG 17-28	.05	.02	.09	.04	.42***		.48***
EEG 6	.01	−.02	.05	.02	.65***	.36***	

Note.—All measures were ranked so that a high score received a high rank.
Adults, upper right; children, lower left.

[a] See Footnote a, Table 1.

[b] Significantly greater than the corresponding coefficient for children at the .01 level.

[c] Significantly greater than the corresponding coefficient for children at the .001 level.

[d] Significantly greater than the corresponding coefficient for children at the .05 level.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

each group which might be termed "somatic." In this latter cluster, the correlations among adults were significantly higher than the corresponding ones among children, though it is possible that the relatively high average rho between HR and RR in children may not have been exceeded among adults had measurement of adult RR been justifiable. The EMG measure in children was only marginally associated with the autonomic measures, while in adults it was fully integrated with them. The two clusters, EEG and somatic, were largely independent within each age group.

Discussion

The positive coefficients of correlation among autonomic and EMG variables within the average adult are of a magnitude previously reported for adults (Hovland & Riesen, 1940; Lazarus et al., 1963; Schnore, 1959). A general activation factor is readily inferable from such a pattern. But what, then, is to be made of the much lower coefficients occurring in the children?

The low intraindividual correlations among somatic variables in children may have some bearing on the question of a communal drive mechanism. Small though the correlations of Table 7 may be, they are still positive; and two thirds of them are significant. The question is why they are so much smaller than the corresponding figures for adults. There is one factor which can serve to lower intraindividual concordance, but cannot serve to increase it. That factor has already been referred to as *situational stereotypy*, i.e., the phenomenon in which different stimulating conditions produce characteristically different patterns of the activity of various bodily responses. Thus, for example, Lacey (1959) reports four situations, labeled "visual attention," "empathic listening," "thinking," and "withstanding pain," the HR and PC responses to these being the measures of interest. For every situation, a majority of his *S*s showed an increase in the PC response; but the HR response varied from predominantly decreasing in the visual-attention condition to predominantly increasing in the withstanding-pain condition. The result is an essentially zero relation between the two autonomic measures.

Applying the example to the present case, the lower concordances of the children might, then, be considered the result of the fact that normal children are more reactive to change in their environment than normal adults are, so that the same "distal" sequence of experimental situations might generate more "proximal" situational change in the child than in the adult.

There is some evidence in the data which supports such a view. Table 6 showed that the range of reaction speed in the average child was far greater than in the average adult. Variability of response may be taken as evidence of variability of stimuli (since under identical stimulus conditions the organism must emit identical responses). Hence, it follows that the children were responding to a larger variety of cues than the adults were. Such reasoning may be applied to the data in a more direct way. Assuming that variability of reaction speed from block to block across the session is a valid index of the number of cues or situations to which *S* was responding, then there should be an inverse relation between range-of-performance scores and strength of concordance within each age group, separately analyzed.

There were 10 adults who had above-average intraindividual rho's for all pairs of the variables HR, PC, and EMG. These 10 "highly concordant" adults had smaller intraindividual ranges of reaction speed than the 11 "low concordant" adults, a difference significant at the .05 level when evaluated by a one-tailed test of the Mann-Whitney statistic *U*. There were 11 children who were highly concordant, having had above-average intraindividual rho's between all pairs of the variables HR, PC, and RR. These 11 children had smaller intraindividual ranges of reaction speed than the 30 low concordant children, and the difference was significant by a one-tailed test of *U* at the .10 level. Such differences, though they are of but marginal significance, provide at least suggestive evidence that situational variability, indexed by intraindividual variability of perceptual-motor behavior, is related to intraindividual physiological concordance.

In view of such a relation, it is possible that the common finding of low but positive intraindividual correlations among somatic measures is evidence primarily of the fact that severe limitations upon situational variability are typically, and of necessity, imposed by laboratory designs. Thus matrices of positive intraindividual correlations are necessary for the inference of a communal drive mechanism, but they will not be sufficient for such an inference until it is demonstrated that the positive correlations among physiological responses exist regardless of the variety of the situations to which *S*s are exposed.

These results on concordance, along with results of the previous section, lead to the conclusion that the conflict or disorganization resulting from responding to multiple cues (e.g., a "divided-set" condition) not only renders perceptual-motor responses independent of physiological responses, but tends to make one physiological response vary independently of another. And EMG data of this section also fit well with those of the previous one: The muscle-tension response in children does not have the same motivational significance that it has in adults.

Adaptation

Sessional Changes

Table 8 presents the data on changes across the session in the response levels of the physiological and overt variables, in the form of the number of *S*s in each age

TABLE 8

NUMBER OF CASES IN EACH AGE GROUP SHOWING DECREASES OR INCREASES IN THE ACTIVITY OF THE VARIOUS RESPONSES OVER THE SESSION

Variable	Adults			Children			*p*, difference of changes for adults versus children
	N decreasing	*N* increasing	*p*	*N* decreasing	*N* increasing	*p*	
HR	16	5	< .05	19	22	*ns*	< .05
PC	21	0	< .001	21	20	*ns*	< .001
Muscle Tension	14	6	*ns*	16	25	*ns*	< .05
EEG 8–12	12	8	*ns*	21	20	*ns*	*ns*
RT	20	0	< .001	20	21	*ns*	< .001
Movement	14	4	< .03	9	23	< .02	< .001

Note.—A "decrease" is a lower score later in the session for HR, PC, EMG, and movement, but a higher score for RT and EEG 8–12. Evaluations of change are by sign test; evaluations of age differences in frequency of change are by chi square; all significance levels are for the 2-tailed test.

*N*s less than 21 for adults and less than 41 for children indicate that some *S*s showed no change.

TABLE 9

AVERAGE CHANGES IN LEVELS OF RESPONSES WITHIN THE NI AND THE I BLOCKS

Measure	Adults				Children				Difference in trend, adults versus children
	Part of block				Part of block				
	1	2	3	$\chi^2 r$	1	2[a]	3	$\chi^2 r$	χ^2
A. Average change within NI blocks									
RT (msec.)	202	195	198	20.2***	354	372	380	24.8***	9.66**
PC (micromhos)	105.8	102.5	100.8	29.8***	186.2		184.7	4.1**	3.41
Muscle Tension (microvolts)	6.7	7.8	7.8	16.2***	7.0		7.1	1.0	7.97**
EEG 8–12 (microvolts)	7.0	7.2	7.3	4.6	10.5		10.5	1.0	1.64
HR (bpm)	77.0	77.8	77.7	2.6	93.7		94.2	7.05***	1.32
B. Average change within I blocks									
RT (msec.)	191	186	190	2.1	341	363	376	22.3***	9.25**
PC (micromhos)	107.2	107.2	105.6	7.7*	191.8		184.7	5.5*	<1.0
Muscle Tension (microvolts)	8.2	10.2	13.4	25.2***	7.6		6.7	5.5*	14.90***
EEG 8–12 (microvolts)	6.7	6.8	6.9	4.2	10.5		10.6	1.0	<1.0
HR (bpm)	80.7	81.5	80.7	3.1	94.7		95.5	2.95	<1.0

Note.—The evaluation of the difference in trend was made by chi square tests on the number of *S*s in each age group who increased or decreased from the first to the last part of the block.

[a] Physiological gradients for children are based upon only two samples within a block.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

group who showed a rise or a fall in the level of response of a variable. The scores used to determine sessional changes were the average level of a response in the first five performance blocks of a session, and the corresponding average level in the last five blocks; each set of five blocks consisted of three NI and two I blocks. There was no evidence of adaptation of any of the responses in the children: Physiological and performance responses stayed at roughly constant levels, and the incidence of restless movement increased. Among the adults, there was strong evidence of adaptation of every response excepting EMG and EEG 8–12; and the adults habituated to a significantly greater degree than the children for every variable except EEG 8–12.

Changes with Blocks of Reactions

Table 9 presents the data concerning the average change in the various responses within the NI (Part A.) and the I (Part B.) blocks of reaction. The entries in the table are means. The analysis of change within a group was done by Friedman's $\chi^2 r$ (Siegel, 1956); the test of age difference in trend was done by chi square, comparing the number of adults and children who changed either up or down in response level from the beginning to the end of a block.

There was a significant within-block adaptation of PC in each age group and under each condition. In performance, however, the groups changed in significantly different ways, the adults having made a net gain in speed as the block progressed, and the children having lost speed. The third notable age difference in gradients of change within blocks occurred in the muscle tension response: adults showed strong increments (negative adaptation) in EMG, and children did not show any change as the series of reactions progressed.

In order to find out whether additional experiences in the laboratory might cause children to show at least the beginnings of adult patterns of change in response levels over time, the records of the 12 children who returned for the brief third session were examined and compared with their second-session records. There was no indication whatsoever that the pattern of response changes varied with an additional session in the laboratory; physiological responses did not decline over the third session, and within-block patterns of change were no different from what they had been in Session II.

Discussion

Over the session, the adults adapted in all responses, as would be expected where a repetitive task is involved. The 6-year-olds, however, failed to habituate in physiological and perceptual-motor responses. In this failure they resemble the groups of psychoneurotic (e.g., Malmo & Wallerstein, 1955) and schizophrenic (e.g., Cohen & Patterson, 1937) adults already mentioned in the introduction.

The lack of adaptation common to children and groups of clinically disturbed adults may have similar origins. The assumption has already been made, in the section on concordances, that the low intraindividual concordance in the children was a function of the relatively large number of different situations to which they were responding. Earlier, evidence was presented in support of the interpretation that the children were in a state of conflict, or divided set, particularly under the NI condition. Their failure to habituate indicates that their susceptibility to the influence of novel and task-irrelevant cues did not diminish as the session progressed. It is just such a persistent response to irrelevant cues to which Shakow (1962) attributes the failure of schizophrenics to habituate in RT and HR. Neither group, children or schizophrenics, seems to develop a "task set" of sufficient strength to exclude what Shakow refers to as the "minor sets" evoked by task-irrelevant internal and external stimulation.

There was some evidence that the children did adapt, if not over the session as a whole, at least within particular periods of it. During the 75 sec. of each block of reactions, they showed a decline both in skin conductance and in reaction speed; but the adaptation of one period never persisted to the next. Each new instruction seems to have been sufficient to dishabituate the children. If adaptation can be taken as evidence that responses are being made to a relatively unchanging set of cues, then the unit of stability of such cues, for the children, was the single block. In contrast, the extensive and continuous adaptation of the adults suggests that they were responding to the entire session as a unit.

One question arising from age differences in adaptation has been ignored until now. It has to do with the course of reaction speed within a block of reactions. The speed of the children simply declined monotonically from the beginning to the end of a series of reactions. Among the adults, reactions were slow at the beginning of a series, then they speeded, and then slowed near the end. Discussion of this difference requires some of the data of the next section and will be postponed until then.

It may be noted, finally, that among adults, the positive slope of the EMG response gradient within the block, and the relatively steeper gradient slope associated with the I blocks, are both phenomena which replicate the results of a number of previous studies (Bartoshuk, 1955; Belanger, 1957; Stennett, 1957b; Surwillo, 1956) in which similar within-trial gradients were found in the responses of both active and passive muscles. The phenomenon has been sufficiently replicable over a number of studies that one may expect with great confidence that the more one motivates an adult to perform, the greater will be his rise in muscle tension from the beginning to the end of his performance trial. Hence the failure to find gradients in muscle tension in the children was the more remarkable and reinforces a point made in the discussion of the results of the previous sections: that the muscle tension response in 6-year-old children does not have the same motivational significance that it has in adults.

Effect of Interstimulus (Preparatory) Interval upon RT

For each *S* the median RT following all occurrences of the 12 long interstimulus intervals was compared with the median RT following all occurrences of the 12 short intervals. Every one of the 21 adults reacted with longer latency (by 18 msec. on the average) following the shorter intervals than following the longer ones; 29 of the 41 children did so (by 14 msec. on the average). Reactions following the short intervals were of longer than average latency 72% of the time, among all of the adults; the corresponding figure for the children was 60%. A chi-square analysis of the number of individuals in each group responding with longer latencies to the short intervals yielded a value of $\chi^2 = 5.86$ ($p < .02$), indicating that the adults were affected by the length of the interstimulus interval to a significantly greater degree than were the children.

Discussion

It is a common finding in studies having adults as *S*s that when preparatory intervals vary around some mean value, reaction speed is slower to the shorter intervals and faster to intervals of mean duration or greater (Drazin, 1961; Karlin, 1959; Klemmer, 1956). The effect is reported to be especially marked among schizophrenic patients (Zahn, Rosenthal, & Shakow, 1963) and geriatric *S*s (Botwinick et al., 1957; Botwinick & Brinley, 1962). The effect seems to depend upon the fact the typical *S* comes to develop some temporal expectancy about when, following a preparatory stimulus, the reaction signal is most likely to occur. He readies his response, apparently, at some interval slightly less than the mean interval following the preparatory signal; and so his temporal set is disconfirmed by the shortest intervals of the series, with relatively long latencies as the result.

Thus the presence of a strong effect of preparatory interval upon RT may be taken to be evidence of some control over the response by an internal, anticipatory, temporal response set. The relatively weak effect in the data of the children may then indicate that control over the response remained primarily exteroceptive: They began their response at the onset of the stimulus, with little or no anticipatory preparation, relative to that shown by the adults. The lack, in children, of a well-developed internalized rhythm response is an interpretation which is supported by two other data already adduced: First, they were unable in the CP condition to keypress in rhythm with the stimulus pace, without prompts; and second, they did not show evidence of the sort of rhythmic breathing common in adults.

Finally, it was noted in the section just previous that the adults began any series of reactions in a block with slow reactions, then speeded, and finally slowed again toward the end of a block. The probable explanation rests upon the fact that the first stimulus of each block had a "preparatory" interval (from the time *E* said "Go") which was both very unpredictable and very short. It probably averaged 2.5 sec., but was not controlled because it depended upon how quickly *E* had stopped the RT mechanism at the end of the previous block. The adults must then have responded to this highly unpredictable initial "foreperiod" with a disorganizing uncertainty; in any case, the initial RT was almost always long. Once the stimulus program was started, however, unpredictability of intervals fell to usual levels, permitting both improved reaction speed and greater relaxation on the part of *S*. For the children, if they had no strong expectancy about the average interstimulus interval, an excessively short one would not have been especially surprising or disorganizing. It seems likely that the children reacted more quickly to the earlier tones simply because they listened best to the early tones. One need only assume that the standard starting instructions evoked strong attentive listening responses early in the block, and that attention grew weaker with time.

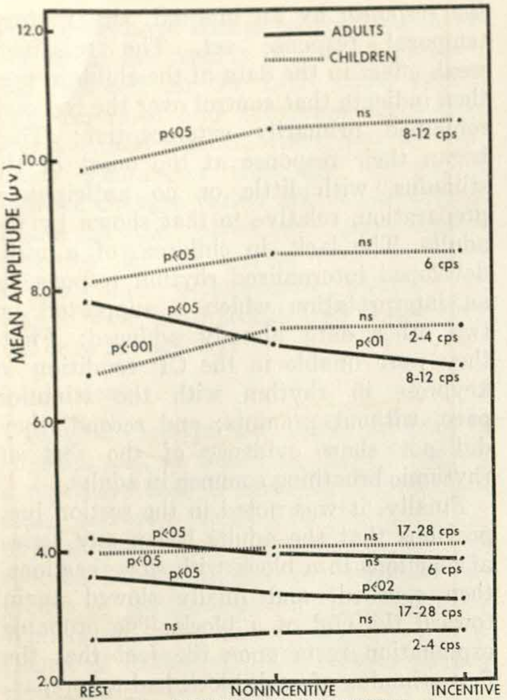


FIG. 9. Average EEG amplitudes in the delta (2–4 cps), theta (6 cps), alpha (8–12 cps) and beta (17–28 cps) ranges, as a function of age and experimental conditions.

Age Differences in EEG Responses

Changes Associated with Experimental Conditions

Figure 9 illustrates the changes in amplitude of the 2–4 cps (delta), 6 cps (theta), 8–12 cps (alpha), and 17–28 cps (beta) activity as a function of conditions for both age groups. Two major phenomena are summarized by the figure. First, as would be expected from the concordance coefficients of Table 7, the different frequencies covary across conditions within each age group. The second and most obvious feature of Figure 9 is the clear interaction between the ages, significant at less than the .01 level for each of the frequencies, in the direction of amplitude change as a function of conditions (see Tables 2 and 3 for the statistics on alpha, which are representative). The interaction is clearest in the comparison between the rest and NI conditions, and it may be recalled that it was between these two conditions that levels of somatic activity rose most sharply in both groups.

Since it has been shown by Stennett (1957a) that alpha amplitudes rise and then fall as PC rises from a low to a moderately high level, it was considered possible that the same effect might have been occurring in the children. To check this possibility the average HRs of each child in each of the three conditions were ordered from low to high, and the pattern of alpha-amplitude changes associated with these rising HRs was determined. Table 10 presents the number and proportion of *S*s in each group falling into each pattern category.

A plurality of the children showed rising alpha amplitudes with increasing HR, but a majority of adults showed decreases in alpha amplitudes with increasing HR. It is this demonstrated nonlinearity of EEG amplitude changes in relation to linear increases of other somatic measures which has made interpretation of EEG data of the previous sections difficult. It has resulted, for example, in great variability of the EEG amplitude change scores of children from the NI to the I conditions, because while some of the children began to show EEG blocking effects concurrently with the rise of autonomic levels in response to the introduction of incentives, other children showed continued increases in EEG amplitude.

TABLE 10

NUMBER AND PERCENTAGE OF INDIVIDUALS SHOWING VARIOUS PATTERNS OF ALPHA AMPLITUDE CHANGE ASSOCIATED WITH INCREASING HRs

Group	Pattern: Monotonic decrease	Monotonic increase	Increase, then decrease	No change, or decrease then increase
Adults				
N	12	2	4	2
%	62	9.5	19	9.5
Children				
N	11	15	8	7
%	26.8	36.6	19.5	17.1

Differences in Frequency and Amplitude

The median dominant alpha frequency of the children was 8.65 cps. Four of them produced frequencies below 8 cps, the lowest being 7.1 cps; 14 others produced frequencies above 9 cps, the highest being 10 cps; and for 4 of the children no dominant frequency could be counted, perhaps because the records of each of them were somewhat "noisy" appearing combinations of very fast and slow activity. Among adults, the median dominant alpha frequency was 10.8 cps; two adults had dominant frequencies less than 10 cps, the lowest being 9 cps; five others had what are called "low voltage, fast" records, with occasional bursts of sinusoidal activity at 14 cps or faster.

The relative contributions of the amplitudes of each of the three measured frequencies to the total measured output was another source of difference between the age groups. As can be inferred from Figure 9, children had relatively, as well as absolutely, more amplitude at 6 cps than adults. There was evidence, however, that the relatively great 6-cps amplitude of the children resulted, to some extent, from the fact that their dominant alpha frequency (8.65 cps) affected the output of the 6-cps resonator to a greater extent than did the dominant alpha (10.8 cps) of the adults. The cutoff characteristics of the filters and resonators were not perfect. However, the outputs of the 2–4 (delta) and the 8–12 (alpha) filters were independent at all voltages. The proportion of total measured activity represented by 2–4 cps activity was .19 for the children, and .11 for the adults. It is clear that the slower frequencies are far more dominant in 6-year-old children than in adults.

EEG Changes Associated with the Key-pressing Response

For both adults and children, reliable changes in the amplitudes of the alpha and beta (but not theta) frequencies occurred in association with the series of time periods before, during, and after the response to the stimulus tone. These changes are illustrated in Figure 10 and appear as the decrease in amplitudes from the S-R period, or the ½-sec. interval just prior to it, to the ½-sec. interval just after it, followed by a rise in amplitudes to prestimulus levels about 1.5 sec. after the response.

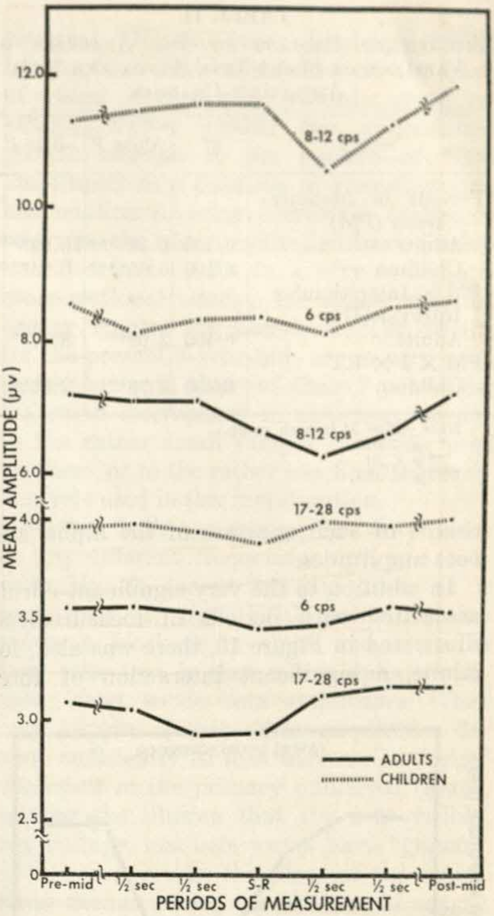


FIG. 10. Average changes in the theta (6 cps), alpha (8–12 cps), and beta (17–28 cps) amplitudes as a function of age and of the periods of measurement from just before to just after the tone-key-press sequence.

The curves of Figure 10 were based upon measured amplitude changes of all selected reactions, regardless of length of foreperiod or speed of reaction. The influence of these latter dimensions of variation were assessed separately for each age group in a 2 × 2 × 7 analysis of variance, consisting of fast and slow RTs, long and short foreperiods, and the seven periods of measurement. Table 11 summarizes the significant

TABLE 11

SIGNIFICANT RESULTS OF THE ANALYSES OF VARIANCE OF SHORT-TERM ALPHA AND BETA AMPLITUDE CHANGES

Source	*df*	Alpha *F*	Beta *F*
Periods of Measurement (PM)			
Adults	6/100	9.58****	15.92****
Children	6/100	6.65****	5.24****
PM × Interstimulus Interval (I)			
Adults	6/100	3.02**	3.17**
PM × I × RT			
Children	6/100	3.79***	2.82**

Note.—$N = 21$ in each group.
** $p < .025$.
*** $p < .01$.
**** $p < .001$.

results of such analyses of the alpha and beta amplitudes.

In addition to the very significant effects associated with periods of measurement, illustrated in Figure 10, there was also, for adults, a significant interaction of foreperiod length by periods of measurement, both for alpha and for beta amplitudes. The basis of this effect is illustrated in Figure 11, which shows changes in beta amplitudes. (The same effects are apparent, though not quite so clearly, in the alpha amplitude changes.) When the foreperiod was long, some decline in amplitude had already taken place, so that the subsequent blocking effect of making the actual response was small; but when the foreperiod was short relatively little preliminary blocking had occurred, and the blocking effect of making the response was relatively large. The dimension of reaction speed, fast versus slow, had no effect on blocking patterns. Among the children, there were no amplitude changes systematically associated with preparatory interval or RT.

The curves for alpha and beta of Figure 10 suggest by their appearance two other effects. The first is that, regardless of frequency, adults seem to show more blocking

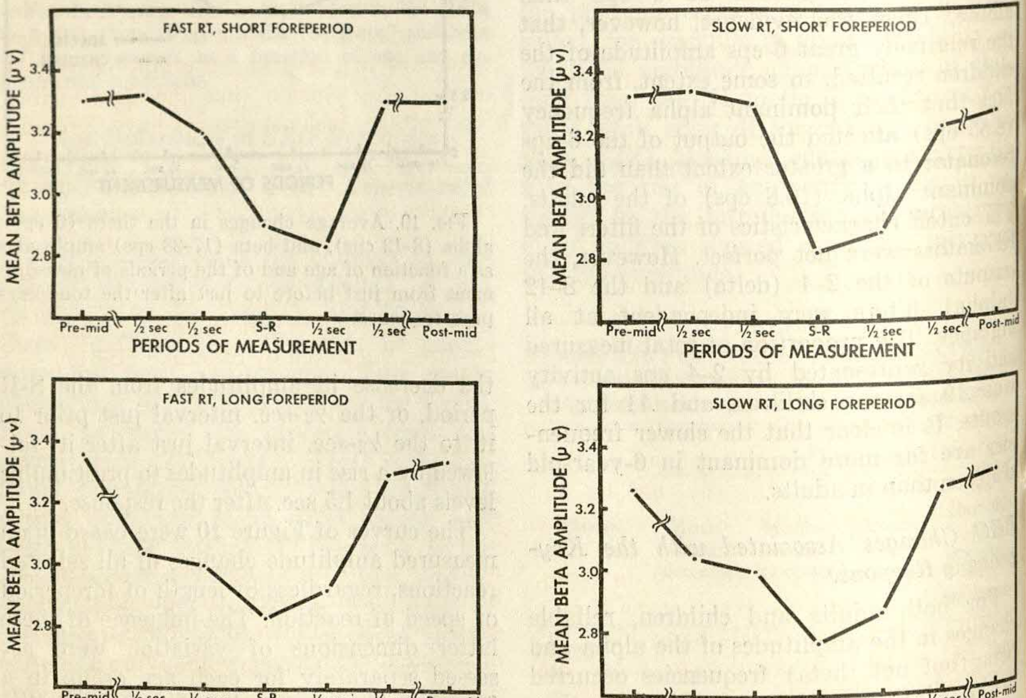


FIG. 11. Average change in the beta (17–28 cps) amplitudes of the adults, as a function of the periods of measurement around the tone-keypress sequence, the length of the foreperiod, and the speed of reaction.

preliminary to the stimulus tone than children; and the second is that, regardless of age, there appears to be more preliminary blocking in beta than there is in alpha.

Discussion

Perhaps the most intriguing question arising from EEG results is this: Why, when a condition of rest is compared with a condition of effort, do children show changes opposite to those of adults in EEG amplitudes, while they show parallel changes in somatic activity? We know, from Stennett (1957a) that a rise in alpha amplitudes will accompany a rise in PC, among normal adults, only when the lowest levels of recorded physiological activity correspond to a state of drowsiness in *S*. And although among the present sample of adults one *S* actually dozed for a few seconds during the final rest period (and showed the atypical increase in alpha with increasing HR), and a few others commented upon the fact that they became progressively relaxed throughout the session, their EEG patterns of change gave little evidence that they had reached the very low state of wakefulness which is accompanied, *in adults*, by diminished EEG amplitudes.

The children, who did show patterns of EEG changes which would indicate drowsiness in adults, were certainly not drowsy in behavior. The incidence of restless moving about; the occassional random keypress and vocalizations; and most of all the obvious "quickness" of the children in response to information, especially the information that they could get out of their chair, convinced *E*, at least, that he had not been dealing with sleepy organisms. One is left with the unexplained fact that giving a 6-year-old something to do (compared with having him "rest"), even if only pressing a key, has effects upon his EEG and PC which are similar to those observed when waking up drowsy adults.

A second age difference in EEG data is better known—namely, the relative dominance of slower activities in the children as compared with the adults. The finding is classical (Henry, 1944; Lindsley, 1939), having been first established on the basis of visual inspection of records; more recently, Walter (1950) has summarized growth changes in the pattern of EEG amplitudes as a function of frequency and has confirmed, using electronic frequency analysis, the older results; so have Corbin and Bickford (1955) in a very systematic cross-sectional study, in which detailed energy spectra are presented. The results for the present 6-year-olds are in very good agreement with those of their 7-year-olds; the small discrepancy in ages may be due to the rather small sample ($N = 5$) used by them, or to the rather less fine frequency analysis used in this investigation.

Within both age groups, the amplitudes at the different frequencies covaried considerably. It cannot be said then, as it sometimes is, that EEG activation involves a blockade in alpha and a rise in beta—the fact is that when alpha amplitudes drop, so do beta amplitudes. What may happen is that alpha amplitudes decline sufficiently so that they are no longer dominant in the primary unfiltered record, leaving the illusion that the now visible, low voltage, fast beta waves have "grown," as it were, to take the place of the alpha. Even though strong correlations of energy changes in different EEG frequencies have been demonstrated, the correlations are far from perfect (see Table 7). It does not seem likely, for example, that beta can be considered simply a harmonic of alpha, as has been suggested by Shminke (1961).

There were similarities between the children and adults in the blocking effects in EEG accompanying response or preparation to respond during the typical tone-keypress sequence. These findings replicate and extend those of MacNeilage (1961) on adults. One of his conclusions was that the beta amplitudes were the quickest, most sensitive EEG indicators of central-nervous-system changes accompanying a simple motor response; and the curves of Figure 10 are in harmony with such a view, for both ages studied. The only indications of age differences in the short-term blocking patterns were that adults manifested patterns which were differentially associated with the length of the preparatory interval, and blocking effects occurred sooner in the tone-keypress sequence. Both of these phenomena appear to constitute evidence of a central-nervous-system correlate of the temporal response set (discussed in the previous section) which made possible the superior reaction speed of the adult.

SUMMARY AND CONCLUSIONS

This investigation had as its purpose the gathering of information pertinent to two areas of interest common to human developmental and physiological psychology. The areas of concern were, first, the nature of the relation between physiological activity and performance in the normal 6-year-old child; and, second, the characteristics of EEG of the child while he is undergoing, with greater or less incentive, successive series of trials in simple auditory reaction time. Corollary questions had to do with the relations of physiological measures among themselves and with their adaptation. A group of young adults provided a comparison set of data, from a design set up within the framework of research in activation and performance.

The adult group manifested the kinds of relation between physiological activity and performance (i.e., a rise in quality with a rise in level of activity) often reported in other similar studies of adults (Duffy, 1962). The children did not. An examination of the nature of the age difference led to the conclusion that it was due primarily to the fact that the children were far less able than the adults to maintain an undivided and persevering attention to their task. The design involved prolonged constraint of movement and delay of reward, two features which probably contributed to the restlessness of the children. Other cases were examined in which adults had produced data similar to those of the children. In such cases, the adults had been either clinically disturbed or had been pharmacologically or experimentally distracted.

Analysis of the data of the adult group failed to support any notion of a causal relation between physiological activity and efficiency of performance, despite impressive covariation of the physiological and perceptual-motor responses. The notion was put forward that covert physiological and overt motor responses are multiple-concurrent responses (Eriksen, 1956) which are functionally related to antecedent conditions, but not to each other. They may covary highly during states of sustained attention to the relevant cues of a repeated task; or they may be poorly correlated in states of distraction, divided set, and the like. The age differences in the physiology-performance relation were explained in terms of the foregoing considerations, especially the notion of set.

There were also age differences in the concordance of physiological measures. Thus, in children, the variation of one somatic measure was far less well correlated with the variation of another within the individual than was the case in adults. This difference was also explained in terms of set, by making the assumption, indirectly supported, that the children underwent multiple sets throughout the course of the experiment. The supporting data in the literature on adults (Lacey, 1959) are that when one deliberately varies the set of *S*, one elicits a pattern of physiological responses unique to each set; the result is that correlations among somatic responses must be attenuated as a function of the number of different sets induced. Since in the present case no experimental effort was made to induce more sets in children than in adults, it was assumed that the children responded to a greater range of the cues available, e.g., that they perceived each new block of trials as a separate task, whereas the adults considered each block as simply a part of a whole task.

The same assumption, concerning the relatively greater variety of cues to which the children were responding, served to account for the fact that the children did not show, as the adults did, adaptation of responses over the session. That is, given that adaptation occurs most strongly with repetition of a task, and is prevented by the introduction of new tasks, the failure of the children to demonstrate habituation has been taken as implying that they perceived each new block of trials as constituting quite a new and different task. In contrast, the children, as well as the adults, did show evidence of adaptation *within* each block—presumably, for the short period taken by a block of trials, no significant change in task was perceived by the children.

The RT of the children was far less dependent than that of the adults upon the length of the preceding interstimulus (preparatory) interval. The result was in harmony with other evidence of the failure of the children to maintain the pace of stimulus presentation. It was interpreted as indicating a far weaker development in the children of a temporal response set which, in the adult, permitted anticipation of the stimulus tone.

The EEG data on the children of this investigation were unusual in that they were taken while *S*s were engaged in sustained, goal-directed performance (i.e., something more than resting, hyperventilating, or opening and closing the eyes). There were two major age differences. The first, confirming a well-established finding, was that the children's EEGs contained relatively far more energy in the slower frequencies than did the adults'. The second finding, hitherto unknown, was that while the EEG amplitudes of adults showed the usual blocking effects with increasing motivation and/or somatic activity, EEG amplitudes of children responded with increases to increasing motivation. These latter EEG responses are similar to those which occur in adults when they are aroused from drowsiness (Stennett, 1957a).

In two important respects also EEG responses of the two age groups were similar. First, there were substantial correlations among the amplitude changes at the different frequencies, so that, for example, a drop in alpha was usually accompanied by a drop in beta. Second, EEG amplitude changes accompanying the making or preparation for the making of the keypress response consisted of a clear dip, or blockade, in the alpha and beta frequencies, with the beta the more sensitive of the two. In the adults, though not the children, the blocking patterns varied as a function of the preparatory interval. This finding paralleled the data on RT as a function of preparatory interval.

Clearly, a considerable reliance has been put on the notion of response set, of maintaining attention to task-relevant cues. It remains to be demonstrated that one could, by setting up the required attentive responses in children, produce patterns of performance and physiological responses that would look more typical of adults. In so far as society, through the schools and the heavier socialization demands occurring with increasing age, sets up greater capacity in children for constraint of movement, longer span of attention, greater tolerance of delay of reward, and the like, one must expect that the present children will produce more adultlike patterns when they are retested at age 8.

There were some points of methodological interest concerning the physiological measures themselves. First, conductance was the most sensitive measure of adaptation; second, HR was the best discriminator of motivational conditions; and finally, EMG of a "passive" muscle is a poor measure of motivation or of general activity level in children.

REFERENCES

AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, **15**, 433–442.

BARTOSHUK, A. K. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, **9**, 215–230.

BARTOSHUK, A. K. Electromyographic reactions to strong auditory stimulation as a function of alpha amplitude. *J. comp. physiol. Psychol.*, 1959, **52**, 540–545.

BARTOSHUK, A. K. Response decrement with repeated elicitation of human neonatal cardiac acceleration to sound. *J. comp. physiol. Psychol.*, 1962. **55**, 9–13.

Belanger, D. "Gradients" musculaires et processus mentaux superieurs. *Canad. J. Psychol.*, 1957, **11,** 113–122.

Belanger, D., & Feldman, S. M. Effects of water deprivation upon heart rate and instrumental activity in the rat. *J. comp. physiol. Psychol.*, 1962, **55,** 220-225.

Bialer, I. Conceptualization of success and failure in mentally retarded and normal children. *J. Pers.*, 1961, **29,** 303–320.

Bindra, D. *Motivation: A systematic reinterpretation.* New York: Ronald, 1959.

Botwinick, J., Brinley, J. F., & Birren, J. E. Set in relation to age. *J. Gerontol.*, 1957, **12,** 300–305.

Church, R. M. The effects of competition on reaction time and palmar skin conductance. *J. abnorm. soc. Psychol.*, 1962, **65,** 32–40.

Cohen, L. H., & Patterson, M. Effect of pain on the heart rate of normal and schizophrenic individuals. *J. gen. Psychol.*, 1937, **17,** 273–279.

Corah, N. L., & Stern, J. A. Stability and adaptation of some measures of electrodermal activity in children. *J. exp. Psychol.*, 1963, **65,** 80–85.

Corbin, H. P. F., & Bickford, R. G. Studies of the electroencephalogram of normal children: Comparison of visual and automatic frequency analyses. *EEG clin. Neurophysiol.*, 1955, **7,** 15–28.

Davis, J. F. *Operator's manual: A. M. I. integrator system.* McGill University, Allan Memorial Institute of Psychiatry, Laboratory for Psychological Studies, 1956. (mimeo)

Davis, J. F. Manual of surface electromyography. *USAF WADC tech. Rep.*, 1959, No. 59-184.

Davis, R. C. Response patterns. *Trans. N. Y. Acad. Sci.*, 1957, **19**(Series II), 731–739.

Drazin, D. H. Effects of foreperiod, foreperiod variability, and probability of stimulus occurrence on simple reaction time. *J. exp. Psychol.*, 1961, **62,** 43–50.

Duffy, Elizabeth. *Activation and behavior.* New York: Wiley, 1962.

Duffy, Elizabeth, & Lacey, O. L. Adaptation in energy mobilization: Changes in general level of palmar skin conductance. *J. exp. Psychol.*, 1946, **36,** 437–452.

Eriksen, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63,** 74–80.

Freeman, G. L. The relationship between performance level and bodily activity level. *J. exp. Psychol.*, 1940, **26,** 602–608.

Haywood, H. C., & Hunt, J. McV. Effects of epinephrine upon novelty preference and arousal. *J. abnorm. soc. Psychol.*, 1963, **67,** 206–213.

Hebb, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62,** 243–254.

Henry, C. E. Electroencephalograms of normal children. *Monogr. Soc. Res. Child Develpm.*, 1944, **9**(Serial No. 39, No. 3).

Hovland, C. I., & Riesen, A. H. Magnitude of galvanic and vasomotor response as a function of stimulus intensity. *J. gen. Psychol.*, 1940, **23,** 103–121.

Jones, H. E. The galvanic skin reflex as related to overt emotional expression. *Amer. J. Psychol.*, 1935, **47,** 241–251.

Karlin, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58,** 185–191.

Klemmer, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51,** 179–184.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein, & M. B. Parloff (Eds.), *Research in psychotherapy.* Washington: American Psychological Association, 1959. Pp. 160–208.

Lacey, J. I., Kagan, J., Lacey, B. C., & Moss, H. A. Situational determinants and behavioral correlates of autonomic response patterns. In P. Knapp (Ed.), *Expression of the emotions in man.* New York: International Universities Press, 1962.

Lazarus, R. S., & McCleary, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58,** 113–122.

Lazarus, R. S., Speisman, J. C., & Mordkoff, A. M. The relation between autonomic indicators of psychological stress: Heart rate and skin conductance. *Psychosom. Med.*, 1963, **25,** 19–30.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Lindsley, D. B. A longitudinal study of the occipital alpha rhythm in normal children: Frequency and amplitude standards. *J. genet. Psychol.*, 1939, **55,** 197–213.

Lucas, W. P., Pryor, H. B., Bost, C., & Pope, S. T., Jr. Growth problems: II. Basal metabolic rate variations in relation to body build, adolescence, and allergy in children. *J. Pediat.*, 1933, **3,** 856–869.

Luria, A. R. *The nature of human conflicts, or, emotion, conflict, and will: An objective study of disorganisation and control of human behaviour.* (Trans. and edited by W. Horsley Gantt) New York: Liveright, 1932.

MacNeilage, P. F. Changes in EEG and other physiological measures during serial mental performance. Unpublished doctoral dissertation, McGill University, 1961.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66,** 367–386.

Malmo, R. B. On central and autonomic nervous system mechanisms in conditioning, learning, and performance. *Canad. J. Psychol.*, 1963, **17,** 1–36.

Malmo, R. B., & Davis, J. F. A monopolar method of measuring palmar conductance. *Amer. J. Psychol.*, 1961, **74,** 106–113.

Malmo, R. B., & Shagass, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11,** 9–24.

Malmo, R. B., & Wallerstein, H. Rigidity and reactive inhibition. *J. abnorm. soc. Psychol.*, 1955, **50,** 346–348.

Mandler, G., Mandler, Jean M., Kremen, I., & Sholiton, R. D. The response to threat: Relations among verbal and physiological indices. *Psychol. Monogr.*, 1961, **75**(No. 9, Whole No. 513).

Mirsky, A. F., & Cardon, P. V., Jr. A comparison of the behavioral and physiological changes accompanying sleep deprivation and chlorpromazine administration in man. *EEG clin. Neurophysiol.*, 1962, **14**, 1–10.

Missiuro, W. Studies on developmental stages of children's reflex activity. *Child Develpm.*, 1963, **34**, 33–42.

Ross, W. R. D., & Davis, J. F. Stable band-pass filters for electroencephalography. *IRE Canad. Convention Rec.*, 1958, No. 860, 202–206.

Schachter, J. Pain, fear, and anger in hypertensives and normotensives: A psychophysiologic study. *Psychosom. Med.*, 1957, **19**, 17–29.

Schnore, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *J. exp. Psychol.*, 1959, **58**, 117–128.

Shakow, D. Segmental set. *Arch. gen. Psychiat.*, 1962, **6**, 1–17.

Shminke, G. A. Alpha and beta waves in the human electroencephalogram. *Pavlov J. Higher Nerv. Activity*, 1961, **11**, 38–41.

Stennett, R. G. The relationship of alpha amplitude to the level of palmar conductance. *EEG clin. Neurophysiol.*, 1957, **9**, 131–138. (a)

Stennett, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61. (b)

Surwillo, W. W. Psychological factors in muscle action potentials: EMG gradients. *J. exp. Psychol.*, 1956, **52**, 263–272.

Telford, C. W. The refractory phase of voluntary and associated responses. *J. exp. Psychol.*, 1931, **14**, 1–36.

Walter, W. G. Normal rhythms—their development, distribution, and significance. In J. D. N. Hill, & G. Parr (Eds.), *Electroencephalography*. London: Macdonald, 1950. Pp. 203–227.

Weiffenbach, J. M. An electrophysiological study of the effect of induced muscle tension on pain thresholds. Unpublished Master's thesis, McGill University, 1962.

Wilcoxon, F. *Some rapid approximate statistical procedures*. Stamford, Conn.: American Cyanamid, 1949.

Wolf, S., & Wolff, H. G. *Human gastric function*. New York: Oxford, 1943.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. New York: Holt, 1954.

Zubek, J. P., & Solberg, Patricia A. *Human development*. New York: McGraw-Hill, 1954.

Zahn, T. P., Rosenthal, D., & Shakow, D. Effects of irregular preparatory intervals on reaction time in schizophrenia. *J. abnorm. soc. Psychol.*, 1963, **67**, 44–52.

(Received March 1, 1964)

Vol. 78, No. 11 Whole No. 588, 1964

Psychological Monographs: General and Applied

PSYCHOTHERAPY EXPERIENCE IN RETROSPECT:

QUESTIONNAIRE SURVEY OF FORMER PATIENTS AND THEIR THERAPISTS[1]

HANS H. STRUPP, MARTIN S. WALLACH, AND MICHAEL WOGAN

School of Medicine, University of North Carolina

This investigation was designed to provide preliminary data on the psychotherapy experience as perceived and evaluated by former patients in retrospect. The major purpose was to gain a better understanding of factors in the patient-therapist relationship that might account for therapeutic change. A comprehensive questionnaire was completed by 44 former patients, paralleled by assessments provided by the patients' therapists. A number of item clusters was isolated from the responses of patients and therapists. There was a substantial consensus between patients and therapists concerning the essential aspects and outcome of the therapeutic experience. Many of the patients' comments pointed to a greater sense of competence in mastering problems in living. While the patient's feeling of having the therapist's respect emerged as an important factor in therapeutic change, the relative importance of the therapist's technical skill remains an open question.

In the mid-1950s, Schjelderup (1955), a Norwegian psychoanalyst, published a study devoted to a careful follow-up of 28 patients who had been seen by the author in psychoanalysis. Despite its many limitations and its failure to live up to the conventional canons of science, of which Schjelderup was keenly aware, the paper commands respect for its down-to-earth honesty, its utter lack of pompousness, and its pioneering spirit. Here is one of those rare attempts to take a dispassionate look at the outcomes of psychotherapy, as objectively as the subject matter permits, without zealousness but also without rancor. In recent years, Pfeffer (1959, 1961, 1963) has conducted intensive follow-up interviews of former analytic patients; and there are several other relevant studies, to which reference will presently be made.

At this stage of development of psychotherapy as a scientific discipline it is not possible to design a "definitive" study which would prove to the tough-minded critic that psychotherapy is effective, but it is altogether feasible to gather limited evidence on the experience of psychotherapy and on the outcome. For this reason it is amazing that so few attempts have been made and are being made, by therapists as well as researchers, to delve into this important problem. On the one hand, there seem to be the carping critics who want to be "shown," by means of a rigorous experimental design, that psychotherapy "does some good." On the other hand, there are the clinicians who scoff at the critics, but for themselves are convinced of the value of their labors. This group appears to be largely resistive to investigation efforts, particularly if they require their active collaboration. Somewhere in the middle is the researcher who feels discouraged by his inability to deliver the definitive study

---

[1] This research was supported by Research Grant MH 02171 from the National Institute of Mental Health, Public Health Service. Grateful acknowledgment is made to the following therapists for their collaboration: D. W. Abse, Mary G. Clarke, J. A. Ewing, D. R. Hawkins, Lucie Jessner, M. A. Lipton, A. J. Prange, J. E. Somers, M. G. Sandifer, H. C. Shands, C. Starling, and C. R. Vernon. The Department of Psychiatry of the University of North Carolina and its former chairman, G. C. Ham, lent the project important support. L. V. Jones contributed valuable statistical advice. The greatest debt, however, is owed to the patients who in a real sense made this investigation possible. Their generosity in filling out a lengthy questionnaire with diligence and frankness is very inadequately acknowledged by this footnote. Unfortunately, they must remain anonymous.

which he feels is demanded of him, but also feels disheartened by the lack of sympathy he receives from the therapists, who might be his allies. The result is that clinicians and researchers go their own ways, some remaining oblivious to the necessity of shedding light on their pursuits, others turning to laboratory studies which ultimately are expected to become relevant to the therapeutic enterprise. Of course, there are exceptions, for example, the long-term project which for almost a decade had been in progress at the Menninger Foundation (Wallerstein, Robbins, Sargent, & Luborsky, 1956), as well as others to be mentioned in the following paragraph. The practice followed by psychoanalytic institutes in the 1920s and 1930s of publishing summaries of treatment results seems to have been discontinued, and a large-scale study undertaken by the American Psychoanalytic Association several years ago has never seen the light of day.

Important contributions have been made by Carl Rogers and his collaborators (Rogers & Dymond, 1954). This work has been invaluable, but its usefulness is restricted because it focuses attention solely on the results of short-term therapy as practiced by one group of practitioners adhering to Rogers' particular theoretical orientation. A number of studies carried out by the Veterans Administration (Lorr, 1962; Lorr, McNair, Michaux, & Raskin, 1962) must also be mentioned. Unfortunately, these publications have had little to say about the nature of the therapeutic experience per se, the kinds of improvement the patients experienced, and the like. The work of Snyder (1961) represents a heroic effort, comparable to Schjelderup's, where a single practitioner and investigator subjects his own therapeutic work to critical analysis. Another major effort in this area has been made by Jerome Frank (Frank, 1961; Frank, Gliedman, Imber, Stone, & Nash, 1959; Stone, Frank, Nash, & Imber, 1961) and his research group at the Henry Phipps Psychiatric Clinic, Baltimore, Maryland. The studies of Feifel and Eells (1963) and Board (1959) have employed methodologies similar to the one used in the present study although their respective patient populations (VA patients and outpatient clinic patients) differ.

These investigations represent the kind of slow and painstaking effort which is a salutary antidote to the prevailing climate which has been sketched in the opening paragraphs.

PRESENT INVESTIGATION

In the present investigation we have set ourselves the task of studying the nature of the psychotherapeutic process *as the patient experienced it,* the personal problems that brought him to psychotherapy, the benefits he felt he derived from his interaction with a psychotherapist over an extended period of time, remaining difficulties in living, and related questions. These accounts will be supplemented by certain parallel evaluations made by the therapist.

The data are entirely retrospective, which obviously is one of the serious limitations of this work. Ideally, one would have liked to obtain evaluations from both patients and therapists at the beginning of therapy, upon termination, and after a specified period of time following psychotherapy. Unfortunately, this was not possible. Therefore, the results incorporate the inevitable distortions to which recall is subject, even if the patients replied with perfect frankness, an assumption which also must be questioned.

A second limitation concerns our inability to make assertions concerning the representativeness of our sample of patients and therapists. All we can say is that the participating therapists were experienced staff members of a major psychiatric treatment center, and they may be considered highly representative of the regular staff. Similarly, the patients were chosen by the therapists from their total patient population, with the help of some criteria (see below). There is reason to believe that the sample is mainly comprised of the kinds of patients usually seen by these clinicians.

We shall forego the luxury of presenting formal hypotheses guiding this investigation because it was apparent from the outset that there were none that could be rigorously tested by the design. Nevertheless, it was our objective to shed light on a number of questions, some of which had been suggested by earlier work done by our group, others derived from our general knowledge of the area and the work of other investigators, and still others which were frankly exploratory.

Of primary interest was the relationship between outcomes (as seen by patients as well as by therapists) and the patients' attitudes toward the therapist and the treatment process. We asked ourselves: To what extent are particular treatment results associated with the way the patient related to the therapist, and the manner in which the therapist related to the patient? What part does the therapist's role behavior (personal style, techniques, etc.) play in treatment outcomes? Can we identify attitude clusters that might differentially describe the treatment atmosphere?

Another concern pertained to the amount of agreement between patient and therapist with regard to their relationship. Do patients and therapists arrive at a consensus about the nature of the therapeutic experience?

There were a number of subsidiary questions which we hoped to explore in the course of this investigation, such as the relationship between length of treatment and outcome, possible differences in therapeutic climate in "psychotherapy" and "psychoanalysis," the bearing which socio-environmental variables might have on the results, and so on. Rather than state these questions formally, we shall take them up in discussing the results.

In addition to the statistical findings which are to be regarded as the hard core of the study, we expected to shed some light on patients' feelings and attitudes as reflected by their responses to open-ended questions included in the questionnaire. It was our expectation that in a number of respects such "free" accounts might prove more valuable and illuminating than answers given to structured items by check marks. However, if combined with the latter, they might facilitate more incisive analyses than might be possible through the use of either technique alone.

Instruments

The major instruments were two questionnaires:[2] the Patient Questionnaire (Questionnaire P), whose final format was decided upon after preliminary tryouts, consisted of 89 items. Of these, 14 were open-ended questions asking for a description of the patient's emotional difficulties before therapy; his satisfaction with the results of therapy; formulation of his major problem(s) in the light of his therapeutic experience; changes experienced; remaining problems and/or new ones; attitude toward the therapist at the beginning of therapy, upon termination, and at the time of answering the questionnaire; description of the therapist's attitude (as the patient experienced it); the kinds of therapist activities to which the patient attributed change; and the like. Twenty-two items were presented in the form of rating scales which attempted to assess the patient's need for further therapy, the amount of benefit derived from therapy, his satisfaction with the results of treatment, feelings about the therapist's level of experience and his age (in relation to the patient's age), the patient's "adjustment" before therapy, the chronicity of his complaints, the degree of his disturbance, anxiety, "internal pressure" before entering therapy, the amount of change experienced as a result of therapy, the degree of change apparent to close and more remote associates, the patient's ability to deal with current problems, disappearance of presenting symptoms, and the like. In addition, 41 items were focused more specifically upon the patient's attitude toward the therapist, the therapeutic process, and his experiences during psychotherapy. This latter group consisted of Likert-type items which the patient rated on a 5-point scale in terms of agreement or disagreement with the stated proposition or attitude. The remaining 12 items dealt with basic biographical data, form, frequency, length of therapy, and related questions.

As will be seen from the foregoing enumeration of items, the questionnaire was designed to provide rather comprehensive coverage of the therapeutic experience, as viewed by the patient in retrospect. Major emphasis was placed on exploration of the therapeutic climate as the patient experienced it;

---

[2] Copies may be obtained without charge from H. H. Strupp, Department of Psychiatry, University of North Carolina, School of Medicine, Chapel Hill, North Carolina or from the American Documentation Institute. Order Document No. 7956 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $2.50 for microfilm or $5.00 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

aspects of therapeutic techniques which might vary from therapist to therapist (as well as for a given therapist with different patients); and attitudes of the therapist, which in part might be fused with his techniques. This focus on the attitudinal and technique components of the relationship was suggested by a series of earlier investigations undertaken by members of our research group (Strupp, 1960; Strupp & Williams, 1960; Wallach & Strupp, 1960) in which the therapist's attitudes toward the patient were shown to be significantly correlated with his clinical assessments, treatment plans and goals, and the manner in which he communicated with the patient.

To obtain ratings paralleling to some extent the patient's self-ratings, a brief rating instrument was designed for use by the therapist (Questionnaire T). The characteristics of this questionnaire and the results obtained from it have been the subject of a previous communication (Strupp, Wallach, Wogan, & Jenkins, 1963). This instrument consisted of 20 items which were to be rated on a 5-point scale ranging from "very little" to "very great." The items included a number of global clinical estimates: defensiveness, anxiety, ego strength, degree of disturbance, capacity for insight, overall adjustment, liking for patient, motivation for therapy, prognosis, countertransference problems, degree to which the therapist enjoys working with this kind of patient, symptomatic improvement, change in basic personality structure, therapist's warmth toward the patient, emotional investment in the patient, patient's warmth toward the therapist, overall success of therapy, quality of the working relationship, patient's satisfaction with the results of therapy, and "depth" of the therapy. A number of the foregoing items were assessed in terms of "before" and "after" therapy, although all the therapists' ratings, like the patients' ratings, were in fact obtained retrospectively.

Therapists

Eleven full-time staff members of the Department of Psychiatry at the University of North Carolina School of Medicine were the psychotherapists participating in this study. With the exception of one PhD psychologist, the remaining therapists held the MD degree. Of the latter group, two were fully trained psychonanalysts (members of the American Psychoanalytic Association), and eight were experienced psychiatrists. The total group had a median of 10 years' full-time experience and devoted a mean of 12.5 hours per week to seeing patients in psychotherapy or psychoanalysis. All of the psychotherapists had had personal psychotherapy, ranging from 200 to 900 hours, with a median of 500 hours. Their mean age was 39 years; median age was 36 at the time of the data collection (1960).

Each therapist was asked to contribute an unselected sample of his former patients. It was stipulated that (*a*) only patients who had been seen in individual psychotherapy or psychoanalysis on an outpatient basis should be included, (*b*) patients should have been adults or late adolescents, (*c*) no effort should be made to select patients in terms of either good or poor treatment outcomes, and (*d*) patients should have had at least 25 hours of psychotherapy.

Procedure

Questionnaires were sent by each therapist to the patients in his sample. No attempt was made to control the therapist's selection of patients except to ask him to adhere to the criteria already mentioned. The therapist was free to compose his own covering letter or to send a suggested standard letter. In either event, letters were typed individually and the therapist signed his name. In order to protect the confidentiality of the patient-therapist relationship no patient was approached directly by the project staff although patients were requested to return the completed questionnaire to the project staff. Patients were not required to sign their names (although many did). They were also given an opportunity to indicate whether they wished their former therapist to see their responses (none of the patients who returned completed questionnaries objected, and most of them seemed to welcome it).

Most participating therapists were reluctant to send follow-up letters urging patients to return completed questionnaires; hence, for the most part, only one letter was written.

Questionnaires were sent to 78 former patients; only two envelopes were returned by the post office as undeliverable. Of the remaining 76, 44 (or 58%) were returned in completed form.

For each patient included in the original sample the therapist completed Questionnaire T, without knowing whether the patient had returned the questionnaire. As patients returned completed questionnaires, the latter were matched with the corresponding Questionnaire T. Beyond this matching procedure, which was essential to the data analysis, no attempt was made to identify individual patients.

Patient Sample

Biographical and Socioeconomic Data

The original sample to whom questionnaires were sent comprised 78 patients, 39 men and 39 women. Of the 44 returned questionnaires, 22 came from men and 22 from women.

In an earlier paper (Strupp et al., 1963) we addressed ourselves to the question of possible differences between patients who return the questionnaire and patients who fail to do so. In order to pursue this inquiry it was necessary, in the case of nonreturnees, to refer to hospital records, clinic charts, and the like, to reconstruct the information which had been provided by the returnees. For various reasons, this search was not entirely successful; and the desired data on some patients proved unobtainable. On the basis of the comparisons which we were able to make, there was a suggestion (statistically not significant at an acceptable level) that returnees had been in treatment longer than nonreturnees (and thus they had been seen for a larger number of interviews), and they had been out of therapy for a shorter period of time than nonreturnees. There were no statistically significant differences with respect to other biographical and socioeconomic variables on which we were able to carry out the required comparisons. This led us to conclude that in these respects the sample of returnees was substantially similar to the nonrespondents. In terms of therapists' evaluations, however, there seemed to be certain systematic differences. Therapists tended to rate returnees as generally having been "better" patients, as having shown greater improvement, and as having established better working relationships with their therapist.

Table 1 summarizes the pertinent biographical information. It will be seen that 70% of the returnees were married, and 61% had at least one child. Median age at the beginning of therapy was 31–32 years and 35 years at the time of responding to the questionnaire. Most patients were fairly well-to-do (median family income between $7,500 and $10,000 per year). Educational level, too, was quite high: the median educational level was 16 years, which places the group at the level of college graduates. Professional people (physicians, psychiatrists, attorneys, etc.) made up 28% of the sample; 16% occupied managerial positions; and 27% were housewives. In terms of Hollingshead's Index of Social Position (Hollingshead & Redlich, 1958, pp. 37–38), the patients seem to belong preponderantly to the upper social classes. The sample is comparable to the clientele of the average psychiatrist or psychoanalyst in private practice.

Form and Length of Psychotherapy

Table 2 summarizes pertinent data concerning the course of therapy. The average patient included in the sample was seen for 166 treatment interviews, and therapy typically lasted 2 years and 4 months. Interviews tended to be on a twice-a-week basis. The mean number of months that had elapsed since termination was 32. While the face-to-face position predominated in interviews, 42% of the group was seen on the couch. Almost three out of four patients had only one therapist, but 14% had been treated by three or more therapists. While the distinction between psychoanalysis and psychotherapy is difficult to draw, it may be presumed (judging from the data on typical frequency of sessions) that approximately one third of the respondents had been in psychoanalysis. This median fee paid by these patients was $15.00 per hour. (Because of possible changes in the fee occurring during therapy, and unreliable recollection on the patients' part, this figure must be regarded as very approximate.)

TABLE 1

Biographical and Socioeconomic Data of the Patient Sample

Variable	*n*
Sex	
Male	22
Female	22
Marital status	
Single	9
Married	31
Divorced	2
Widowed	2
Number of children	
None	17
One	7
Two	12
Three	7
Four	1
Occupation	
Professional	15
Housewife	15
Managerial	6
Technical	4
Clerical	4
Gross family income	
Less than $5,000 per year	6
$5,000–$7,500	8
7,500–10,000	10
10,000–12,500	6
12,500–15,000	3
15,000–20,000	4
More than $20,000	5
Blank	2
Age (at present)	
20–25	4
26–30	7
31–35	11
36–40	8
41–45	6
46–50	3
51–55	2
56–60	2
61–65	1
Education	
11 years	2
12 years (high school)	8
13 years	2
14 years	2
15 years	1
16 years	10
17 years	1
18 years	8
19 years	2
20 years	6
More than 20 years	2

The large percentage of hospitalizations for psychiatric reasons is somewhat unusual: fully 45% of the group for whom data were available had been patients in psychiatric hospitals. The

TABLE 2

Data Concerning the Course of Therapy

Variable	*n*	*M*
Frequency of sessions per week		
Less than 1 hour	3	
1 hour	15	
2 hours	11	6.4 (per month)
3 hours	10	
4–5 hours	4	
Blank	1	
Length of therapy		
0–25 hours	4	
26–99 hours	9	
100–199 hours	12	166
200–299 hours	10	
300–499 hours	7	
Blank	2	
Interview position		
Face-to-face	25	
Couch	18	
Blank	1	
Number of months in treatment		28.4
Number of months since termination		32.0
Number of therapists		
One	31	
Two	7	
Three or more	6	
Previous psychiatric hospitalization[a]		
Hospitalized	15	
Nonhospitalized	18	

[a] This information was gathered from the patients' hospital charts after the close of the study. In a number of instances these records were not obtainable.

reasons for this finding are not entirely clear, but may be partly accounted for by the fact that the therapists represented in this sample were all staff members at a university hospital which included a psychiatric inpatient unit. Thus, they may have hospitalized their patients more readily than psychiatrists whose professional work is restricted to office practice; and conversely, they may have accepted a disproportionate number of inpatients for subsequent out-patient treatment.

Results

Statistical Treatment of the Data

Questionnaire responses provided by the patient sample fell into two broad groups: responses to rating scales and precoded items, which were readily quantifiable, and responses to open-ended questions, which were not. For the first group of item responses, frequency distributions were obtained after the data had been recorded on IBM cards, and a table of intercorrelations was prepared. Responses to open-ended questions were transcribed verbatim, with the expectation that some categorization of responses to individual items would be made eventually. In addition, a table of intercorrelations was obtained including both patients' and therapists' ratings.

In dealing with the responses to Questionnaire P, it was our objective to identify as many meaningful dimensions as could be distinguished, to study the relationships of these dimensions to each other, and to explore the extent to which they might relate to parallel assessments made by each patient's therapist. In this way it was hoped to learn whether patients' feelings, attitudes, and self-evaluations were wholly idiosyncratic or whether there was agreement between patients and therapists, both with respect to their attitudes toward each other and the outcome of therapy. The technique of factor analysis was considered unfeasible for this purpose because of the large number of items involved (which exceeded considerably the number of patients responding to the questionnaire), the lack of independence of item responses, and the like. Instead, we adopted a less rigorous approach, an informal cluster analysis, which is believed to have yielded results substantially identical to what would have been the outcome of a factor analysis.

In deciding on the optimum number of item clusters, the project staff, working both individually and as a group, inspected the patterns of intercorrelations (separately for Questionnaire P and for Questionnaire T) and grouped items which showed substantial correlations with each other and low correlations with other items. This analysis yielded 13 clusters for Questionnaire P and 5 clusters for Questionnaire T. In each instance, a small number of items were considered as not belonging to any of the clusters, and they were retained to be treated singly in subsequent analyses. A number of items in Questionnaire P, on the basis of extremely high correlations with other items, were judged to be largely synonymous with other items already represented in a given cluster (redundant), and were consequently eliminated. No item was included in more than one cluster (see Footnote 2).

In subsequent discussions we shall adhere to the following distinctions:

1. Clusters: this refers to the item composites referred to above.

2. Items retained separately: for the reasons given above, some items were retained separately throughout the data analysis.

3. Questionnaire items—other: sometimes we shall consider questionnaire items as separate units. Such references may pertain to items included in the clusters (1), items retained separately (2), and occasionally to items which were omitted from other analyses.

Since the various item clusters will be referred to frequently throughout this report, it may be helpful to present a brief characterization of each in terms of what we conceive to be its essential meaning. It should be pointed out that the clusters were derived empirically (on the basis of the magnitude of the item intercorrelations). Hence it may occasionally appear as if an item—on the basis of semantic considerations—does not "belong."

Therapists' Assessments.

Cluster T_1: Overall success of therapy (3 items). A global assessment of the amount of symptomatic improvement and success of therapy.

Cluster T_2: Remaining disturbance (5 items). A composite of several clinical assessments which are believed to measure the patient's general emotional adjustment following psychotherapy.

Cluster T_3: Warmth of the relationship (7 items). Represents an estimate of the therapist's liking for the patient, the degree to which he enjoys working with "this kind" of patient, the extent of the therapist's emotional investment in the treatment, and the amount of warmth he felt toward the patient. On the basis of the empirical findings, the therapist's estimate of the prognosis "before therapy" was judged to be part of this composite.

Cluster T_4: Patient's capacity for intensive therapy (4 items). This cluster consists of the therapist's estimate of the patient's capacity for therapy, a rating of the intensity of therapy, and an estimate of the degree to which a basic personality change occurred in the patient.

Cluster T_5: Adjustment before therapy (4 items). This composite, perhaps to some extent a counterpart of Cluster T_2 (Remaining disturbance) includes estimates of anxiety, ego strength, and overall adjustment—all estimated before therapy.

Patients' Assessments.

Cluster P_1: Therapist's warmth (9 items). This cluster appears to represent a global estimate, by the patient, of the therapist's attitude during therapy, as the patient recalled it. A number of the items are perhaps fairly direct indicators of the therapist's professional attitude (e.g., stiff and formal, natural and unstudied, extremely passive), but they shade into characterizations of the kind of person the patient considered the therapist to be (e.g., cold and distant, showing a good deal of warmth). One item dealt with the patient's trust in the therapist's integrity.

Cluster P_2: Amount of change (6 items). Represents a global estimate of the amount of benefit the patient felt he derived from therapy, the amount of change he felt became apparent to others, the extent of symptomatic improvement, and his overall satisfaction with the therapy experience.

Cluster P_3: Present adjustment (4 items). This composite, which appears to be made up of closely related items, attempts to measure the degree to which the patient currently feels in need of further therapy, the adequacy of his current adjustment, and his ability to cope with current problems.

Cluster P_4: Amount of change apparent to others (2 items). Combines two items which asked about the amount of change apparent to close friends on the one hand and co-workers and acquaintances on the other.

Cluster P_5: Therapist's respect and interest (3 items). Deals with the extent to which the patient felt respected, the therapist's perceived interest in helping, and the degree to which the patient felt accepted as a worthwhile person.

Cluster P_6: Uncertainty about therapist's feelings (5 items). This cluster is essentially the obverse of Cluster P_5 (Therapist's respect and interest). Feelings of being "just another patient," not having one's real feelings understood by the therapist, being doubtful about the therapist's "real" feelings and his interest in helping seem to characterize this composite.

Cluster P_7: Intensity of emotional experience (4 items). Gauges the extent to which the patient experienced psychotherapy as an intensely emotional experience in which transference feelings were mobilized.

Cluster P_8: Use of technical terms by the therapist (3 items). Focuses on the extent to which the therapist used technical terms and emphasized relationships in the patient's current life. Frequent references to psychoanalytic theory completes this triad.

Cluster P_9: Degree of disturbance before therapy (4 items). Attempts to assess the patient's felt anxiety, the degree of his disturbance, the extent to which he felt impelled to do something about his emotional problems, and his general adjustment—all of these variables referring to the beginning of therapy, as seen by the patient in retrospect.

Cluster P_{11}: Intensity of treatment (3 items). This cluster is composed of factual items dealing with the frequency of sessions per week, the length of therapy, and face-to-face contact versus use of the couch. This cluster is treated as a relatively objective index of the intensity of therapy.

Cluster P_{12}: Importance of emotion in therapy (2 items). These items attempt to gauge the therapist's relative emphasis on the importance of the emotional experience versus intellectual understanding. It is assumed that in intensive psychotherapy the former is stressed more than the latter.

Cluster P_{13}: General dissatisfaction (3 items). Frequent change of therapists, a long period before marked change occurred, and criticisms of the questionnaire appear to point to a general dissatisfaction with the therapy experience, as seen by the patient in retrospect.

As a first step in analyzing our results, cluster scores were computed by summing appropriate item responses. These cluster scores and separate items for Questionnaire P as well as for Questionnaire T (33 variables in all) were then intercorrelated. This set of intercorrelations formed the basis for most of the quantitative findings to be reported (see Footnote 2). Throughout, correlations reaching or exceeding the .05 level (two-tailed test) were treated as statistically significant.

The general plan for discussing the results is as follows: We shall first present data pertaining to the nature of the patients' problems, the reasons for seeking psychotherapy, the chronicity of their emotional problems, and the like. Then we shall turn to the patients' description of the therapeutic experience itself, their attitudes toward the therapist, their feelings during therapy, etc. The next step will entail consideration of the statistical findings, particularly the relationships between clusters in Questionnaire P. After turning to the therapist's view of the patient, we shall relate the therapists' evaluations to those by the patients.

A. Patient's View

Presenting Symptoms

We did not obtain formal diagnoses from the treating therapists. This omission was largely determined by our judgment that the usual short-hand labels would be of little value to the central purpose of the study—the small size of the sample would not permit meaningful breakdowns by diagnostic categories anyway—and, considering the necessity of limiting the demands on the therapists' time, we felt that other kinds of ratings were potentially more important and meaningful.

As might be expected, the patients presented a wide variety of neurotic symptoms which typically appeared in combinations.

Table 3 summarizes the complaints as the patients recalled them. It is evident that the difficulties which brought these patients to psychiatric treatment are in no way unusual; in fact, they may be considered rather typical of the kinds of problems which the average practitioner encounters in his daily work. No useful purpose would be served by efforts to arrive at dynamic formulations from these listings. Instead we shall restrict ourselves to a few observations, realizing that the patients' descriptions are undoubtedly incomplete and subject to some degree of distortion.

Virtually all patients who responded to the question listed multiple symptoms and problems. Most prominent among these were: anxiety in its various forms, unhappiness and other kinds of dissatisfactions, physical symptoms (it seemed pointless to determine how many were of psychosomatic origin), and depression. Difficulties in interpersonal relationships, vocational problems, and phobias were also mentioned with some degree of frequency. By and large, it may be said that the patients represented in the sample were suffering from neurotic conditions with the common "mixed" symptomatologies.

TABLE 3

SUMMARY OF SYMPTOMS AND COMPLAINTS LEADING PATIENTS TO ENTER PSYCHOTHERAPY

(N = 41[a])

Symptom or complaint	n	%[b]
Depression	12	29.3
Anxiety, nervousness, tension	21	51.2
Phobias	6	14.6
Obsessive-compulsive	2	4.9
Acting out, impulsiveness, alcoholism	5	12.2
Physical symptoms, psychosomatic problems	14	34.1
Conversion reactions	3	7.3
Paranoid concerns	2	4.9
Sexual problems	3	7.3
Interpersonal relationships	10	24.4
Work inhibition, vocational problems	7	17.1
Feelings of inferiority, unhappiness, failure, lacking self-esteem, indecision, guilt	18	43.9
Total	103	—

[a] Three patients failed to supply pertinent data.
[b] Percentages exceed 100% because of multiple symptoms.

Degree of Disturbance

Several questions were directed at the severity of the disorder, as the patient experienced it. Responses to Items 69, 70, and 71 clearly indicate that virtually all patients experienced fairly intense anxiety and found themselves under considerable strain and tension which propelled them to seek professional help. The distribution of Item 69 ("How severely disturbed were you?") shows somewhat greater dispersion than the other two items, but again the vast majority of patients considered themselves to be at least "moderately disturbed" prior to entering therapy. Anxiety, evidently, was the strongest motivating force in seeking help.

69. How severely disturbed did you consider yourself at the beginning of your therapy?

	n	%
Extremely disturbed	7	15.9
Very much disturbed	21	47.7
Moderately disturbed	10	22.7
Somewhat disturbed	4	9.1
Very slightly disturbed	2	4.6

70. How much anxiety did you feel at the time you started therapy?

	n	%
A tremendous amount	13	29.5
A great deal	26	59.1
A fair amount	5	11.4
Very little	0	0
None at all	0	0

71. How great was the internal "pressure" to do something about these problems when you entered psychotherapy?

	n	%
Extremely great	17	38.6
Very great	11	25.0
Fairly great	11	25.0
Relatively small	4	9.1
Very small	0	0
Extremely small	0	0
Blank	1	2.3

A somewhat more general question attempted to gauge how well the patient felt he was getting along before therapy (Item 67). The responses to this item indicate a somewhat more favorable picture, suggesting that despite inner discomfort and emotional disturbance, patients were able to function to some extent, i.e., they did not view themselves as totally incapacitated. Nevertheless, most of them regarded themselves as quite ill.

67. How well did you feel you were getting along then?

	n	%
Very well	0	0
Fairly well	4	9.1
Neither well nor poorly	5	11.4
Fairly poorly	10	22.7
Very poorly	15	34.1
Extremely poorly	10	22.7

Chronicity

Item 68 inquired concerning the length of time before entering psychotherapy that the patient felt in need of professional help. Almost one half of the patient group described their difficulties as being of relatively recent origin, very few giving recognition to the probability that their problems were of long standing. At any rate, it seems likely that exacerbations in the recent past accounted for a large proportion of the patients taking the decisive step of entering psychotherapy.

68. How long before entering therapy did you feel in need of professional help?

	n	%
Less than 1 year	21	42.7
1–2 years	15	34.1
3–4 years	3	6.8
5–10 years	4	9.1
11–15 years	1	2.3
16–20 years	0	0

Reasons for Termination

The data indicate that 43% of the respondents terminated therapy by mutual agreement with the therapist. From this answer we do not know, of course, whether the therapist's or the patient's feelings in the matter were decisive or whether the decision was in fact "mutual." In 25% of the cases, external factors apparently played an important part in bringing therapy to a close, and 20% of the patients felt that it was their own decision. It is somewhat surprising that in only 4.5% of the cases was the therapist credited with having suggested termination. This might mean that therapists by and large were reluctant to make direct recommendations which the patient remembered as such. It is also possible that they gently steered the patient toward termination, but proceeded in such a way that the patient experienced it either as his decision or as one mutually arrived at. However, in relatively few instances did the therapist set specific time limits.

17. What led to the termination of your therapy?

	n	%
My decision	9	20.5
My therapist's decision	2	4.5
Mutual agreement	19	43.2
External factors	11	25.0
Other	1	2.3
Blank	2	4.5

Assessments of Therapeutic Results

The questionnaire included a number of items dealing with the results of therapy, as the patient experienced them. The reason for presenting these questions of highly similar meaning derived from the anticipation that overall assessments would be permeated by a large halo effect, and that slightly different wordings of items might produce in some (unforeseeable) instances better differentiation among the respondents.

Item 18 dealt with the amount of benefit the patient derived from therapy. It will be seen that over two thirds of the respondents described the benefit as having been substantial, with only a small proportion of the sample voicing negative opinions. Yet, approximately 18% of the group were guarded in their evaluations.

18. How much have you benefited from your therapy?

	n	%
A great deal	30	68.2
A fair amount	6	13.6
To some extent	6	13.6
Very little	1	2.3
Not at all	1	2.3

Item 19, devoted to the patient's satisfaction with the treatment results, while providing for a larger number of response categories, discloses a similar distribution.

Fifty-five percent were extremely satisfied or highly satisfied with the results; another 23% were "moderately satisfied," with approximately 20% having various degrees of ambivalence. It should be noted, however, that only 4.6% actually committed themselves to the statement "dissatisfied."

19. Everything considered, how satisfied are you with the results of your psychotherapeutic experience?

	n	%
Extremely dissatisfied	0	0
Moderately dissatisfied	1	2.3
Fairly dissatisfied	1	2.3
Fairly satisfied	7	15.9
Moderately satisfied	10	22.7
Highly satisfied	12	27.2
Extremely satisfied	12	27.3
Blank	1	2.3

Item 79 attempted to probe somewhat more deeply by inquiring into the amount of symptomatic improvement rather than overall improvement. The response distribution shows that less than 10% of the patients were prepared to assert that their symptoms or complaints which brought them to therapy had "completely disappeared." On the other hand, by applying a somewhat less stringent criterion, fully three-quarters were willing to assert that there had been at least "considerable" improvement. Slight improvement was conceded by 20%, but 4.5% noted no improvement at all.

79. To what extent have your complaints or symptoms that brought you to therapy changed as a result of treatment?

	n	%
Completely disappeared	4	9.1
Very greatly improved	17	38.6
Considerably improved	12	27.3
Somewhat improved	9	20.5
Not at all	2	4.5
Got worse	0	0

A somewhat more coarse-grained question focused on the amount of change the patient experienced as a result of psychotherapy (Item 73). Again, a substantial 70.5% of the group noted "a fair amount" or "a great deal" of change; 18.2% admitted having changed "somewhat"; 9.1% had changed "very little" or "not at all." We shall soon examine more closely the kinds of changes the patients reported.

73. How much do you feel you have changed as a result of psychotherapy?

	n	%
A great deal	19	43.2
A fair amount	12	27.3
Somewhat	8	18.2
Very little	3	6.8
Not at all	1	2.3
Blank	1	2.2

It seemed important to explore, in gross fashion, the amount of time required for therapeutic change to be noted by the patient. It is a common observation that therapeutic gains are achieved very gradually, occasional dramatic improvements in rare patients notwithstanding. Responses to Item 80 tend to substantiate this notion. Slightly more than a third of the patients experienced marked change within 6 months after the beginning of therapy, and an additional 18% felt improved within a year.

80. How soon after entering therapy did you feel any marked change?

	n	%
Less than 6 months	17	38.6
6–11 months	8	18.2
1 year–1½ years	5	11.4
19 months–2 years	3	6.8
More than 2 years	2	4.5
Blank	9	20.5

But, for about 25% of the group it took considerably longer before they felt better. It is noteworthy that fully one fifth of the group failed to answer this item.

Another question (Item 88) dealing with the patient's belief in the value of psychotherapy was phrased in terms of the degree to which he might recommend this method of treatment to a friend with emotional problems. Fully three-quarters unequivocally would recommend psychotherapy, whereas the remainder would recommend it but "with reservations." While this item yielded a much poorer distribution than the preceding ones that have been commented upon, it appears that in all instances responses to questions dealing with improvement, change, symptom relief, or satisfaction disclosed that approximately two thirds of the patients responding to the questionnaire reported positive results. This percentage agrees rather well with typical studies in the literature in which the *therapist's* evaluation, however, is usually taken as the criterion. For a summary of pertinent studies see Eysenck (1961) and Malan (1963).

88. How strongly would you recommend psychotherapy to a close friend with emotional problems?

	n	%
Would strongly recommend it	33	75.0
Would mildly recommend it	0	0
Would recommend it but with reservations	11	25.0
Would not recommend it	0	0
Would advise against it	0	0

Reasons for the patient's satisfaction (or dissatisfaction) with psychotherapy were elicited by means of an open-ended item. The gamut of responses is best exemplified by a number of verbatim quotations from the questionnaires:

> Although psychosomatic symptoms have been relieved, the basic apathy and depression have improved only moderately.
>
> Helped me much more than I expected it to before starting.
>
> My outward "trouble" was and is overweight. I am still overweight, but I find that I am getting along better with people and can help my wife, my parents, my students, and myself as a result of my psychotherapy.
>
> There has been little change in my response to stressful situations. My ability to cope with problems has not changed. The great change was the completion of medical school with the great reduction in stressful situations which followed.
>
> Have become convinced of validity of certain basic responses. Therapy has been an *experience* which has conditioned me to view myself and my actions with the tools of psychiatry—that is, within a consistent framework. Have come to view myself more objectively, less remorsefully and more forgivingly.
>
> The extreme anxiety which caused me to seek treatment has been practically eliminated, and I find I am much better able to handle normal tensions and anxieties.
>
> The final conclusion of therapy (other than that therapy goes on forever—which it was good for me to see) was that in things which I could not feel, I would have to discipline myself, to carry thru, this has been effective in many things; still in several things, especially sex, I need to clear up my thinking and emotional feeling more—discipline alone is not enough and is not satisfying to the extent that I feel I might find satisfaction.
>
> I think I obtained a certain amount of insight into some of my problems. However, they were not worked out, and I suspect they could not have been worked out in the time available.
>
> Heavens above—I could write a book! My whole life has changed and I now live comfortably with myself and my family. Not only this, I have had a baby—now eight months old—and this has brought us all a joy beyond words.
>
> Life is so much more rewarding. This experience is something good to pass on to those who are dearest and nearest to me.
>
> I don't have the spells of fear and anxiety—though at times, I have a little, but not often.
>
> I am pleased with progress but feel I have as yet unsolved problems.
>
> Great increase in emotional tone and ability to give vent to uncomfortable feelings when they arise. Thus, get a big kick out of a sunset or a pair of well-shaped female legs. Can express all ranges of feelings and react to everything (pleasant and unpleasant). Personal insight and ability to analyze my own problems as they arise. Vastly improved ability to understand the needs, problems, motivations, and feelings of others.
>
> I still have same symptoms as at start, but less severe anxiety. I am in a less taxing situation now—but cannot face the same situation—school—in which I found it necessary to seek therapy.

These quotations bear out the general impression given by the distributions of responses previously cited. They also support the well-known clinical observation that while patients often enter therapy with a few specific symptomatic complaints for which they wish relief, as therapy progresses, they begin to perceive a wider variety of situations and behaviors which are contradictory, incompatible with more adult purposes, and the like. Often, as these difficulties come to light and begin to be resolved, the patient loses his concern with the problems which originally brought him to therapy. The significance to the patient of the resolution of such "deeper" difficulties may be seen in the fact that even when patients leave therapy *without* experiencing marked symptomatic relief, they frequently feel that they have greatly benefited from treatment.

Patients' Evaluations of Personality Changes

Item 74 presented the respondent with a rather difficult task:

> Describe the most important changes you have experienced—internal as well as external. (Try

to list in order of importance and be as specific as possible. Also try to indicate which changes have been lasting and which ones, if any, have not been.)

The difficulties were in asking the patient to be a clinician, or at least a reasonably objective observer of himself. While some patients tried to comply with the instructions of listing changes in order of importance and differentiating between lasting and more transitory ones, most of them made no pretense at being able to make such distinctions. Nevertheless, 39 of the 44 respondents made an effort to describe comprehensively changes in feelings, attitudes, and behaviors; and they did so in remarkably forthright, self-revealing, and self-critical fashion. While the many shades of meaning virtually defied categorization, an attempt was made at a coarse grouping. Beyond these classifications, the patients' answers to this item, a sample of which we shall quote verbatim, are considered among the most valuable data emerging from this investigation. Their value, as we see it, lies in the import as human documents which attest to personality and characterological changes resulting from (or at least attributed to) the experience of psychotherapy. In terms of the patients' descriptions, changes effected by psychotherapy were multifaceted indeed, permeating every aspect of life. Parenthetically, this illustrates the futility of labels like "improved" which usually obscure more than they reveal. It will be noted that many changes, as seen by the patients, appear to be qualitative as well as quantitative. They reveal a markedly different outlook on life, modifications in interpersonal relations, etc.; and lend themselves to the inference that in a fair number of instances a far-reaching reorganization of the personality structure has occurred. In other words, beyond a diminution of the intensity in the symptoms which originally brought the patient to psychotherapy, the matrix of the personality structure in which the neurotic symptoms were imbedded seems to have changed.

TABLE 4

SUMMARY OF CHANGES RESULTING FROM PSYCHOTHERAPY REPORTED BY PATIENTS

($N = 39$)

Category	n	%[*]
Greater awareness of feelings and impulses, more realistic appraisal of own needs, greater independence, flexibility, insight into own past, objectivity, sense of responsibility, more enjoyment and direct satisfactions	28	71.8
Increase in self-esteem, self-respect, self-acceptance; greater sense of well-being, better "integration"	15	38.5
Better interpersonal relationships, improvement in sexual relations, greater tolerance of others	11	28.2
Improvement in presenting neurotic symptoms, lessening of anxiety, greater ability to deal with symptoms and problems	21	53.8
More active in working out solutions, less passive and fatalistic, more energy, more enjoyment from work and activity	7	17.9
No change, changes due to factors other than psychotherapy, persistence of difficulties (sometimes coupled with improvements in other areas)	7	17.0

[*] Exceeds 100% because of multiple classifications.

Table 4 represents an attempt to group the patients' responses. Obviously, there is a good deal of overlap among the categories, and no claim is advanced concerning their separateness or distinctness. Yet, for purposes of exposition it may prove useful to discuss changes in each of the classes.

Greater Awareness of Feelings and Impulses

A majority of the patients made comments which appear to be classifiable under this heading. While many nuances are discernible in the patients' comments, the central theme appears to be *insight into one's feelings and motives*. In the course of therapy most respondents seem to have achieved some measure of what traditionally has been postulated as the principal desideratum in psychoanalysis and intensive psychotherapy—greater self-awareness and emotional insight. From this achievement, as will be seen from the verbatim quotations, seem to flow a wide variety of beneficial changes in the perception of reality, acceptance of oneself, the manner of relating to others, style of life and life goals, techniques of solving interpersonal problems and intrapsychic responses to one's feelings and impulses (see next heading). Altogether the patients seem to have achieved an improved sense of mastery; concomitant with disappearance or reduction of symptoms; a greater sense of being alive; a clearer, more plastic view of themselves; a greater ability to give fully of themselves; and a greater *joie de vivre*. Metapsychological analyses of such changes have been attempted by many psychoanalytic writers; they appear to be largely congruent with the patients' subjective accounts.

Typical comments by the patients include:

More awareness of what my feelings are—admitting aggressive as well as sexual motivation. Seeing more similarity between myself and other people. Understanding more about by relations with my parents. Feeling calmer and more secure. Understanding my wife better. The above changes seem to be lasting, at least as far as I can tell so far.

As to internal changes, I have had more respect for my own ideas and abilities and I have been much more successful in "getting close" to people. My wife and I are extremely close together. This has been my most important change. My adjustment to her and my ability to get along with other people and to accept them on their own terms.

Ability to get along with others better; ability to take criticism better; less critical of others; more understanding of others' problems and needs; more patient with students (and others); less concern about what others think; more willing to rely on my judgment; better able to accept my limitations; more objective as I face problems.

(1) A view of myself as an organic whole changing gradually thru time rather than as a series of disjointed snapshots having no integral relation to each other. (2) A reduction in feelings of guilt about not being properly affectionate toward members of my immediate family. (3) Intellectual insight with certain rather intense hostile feelings towards lots of persons having a safely peripheral role in my life. (4) Growth of the attitude that it was not necessarily entirely unreasonable to consider myself a member of the class of men (as opposed to human male adults). (5) Growth of attitude that, if I wished, I could be an agent for change in certain situations. (6) The awareness that I was lonely.

(1) I do not get sick as much as I did and when I do the intensity is not as great. (2) I have more confidence in myself in how I look, talk, etc. (3) I am not afraid of the opposite sex and I associate with them now. (4) I have learned to drive and dance, without therapy I don't think I would have. (5) I think therapy has matured me. (6) I don't have to play horses or masturbate. To sum it up, I look at myself like a man and not like a little boy.

Many fears have completely disappeared (i.e., fear of mental illness, being alone) and in turn there are more times when I feel "cool, calm and collected" rather than like a "cat on a hot tin roof." Periods of anxiety still "catch" me but I usually feel that they will pass and bear with them. The most helpful part to me I think is that I no longer try so desperately to hide hostility from myself—I usually try to accept it, let it go and move on—I have found that I have better family relationship, but I still have a long way to go. Finally I do like myself more than I did in the beginning, and it is not as important to "will" myself to conform.

The most important change has been a great increase in my ability to work through problems and a certain determination to do so. There seems to be a more realistic appraisal of people, both pro and con. These are permanent changes. Also, I would include what I would name the ability to gamble, to take chances, to wait for a second wind. I am able to maintain longer relationships with men in a wholesome framework. A change manifested in therapy, but not a lasting one, was that I took my problems to one person and did not feel the need to talk to friends, etc. For about a year after therapy, I felt the therapist's presence with me internalized, but the effects have about worn off, and I have to relieve myself by talking to friends again. This brings some confusion.

I am breaking the habit of trying to please and "do for" my parents first. All my life seems to have been a series of shouldering their problems, not to cause them worry or to hurt them. I've never tried to be a personality of my own being, but rather what they wanted. When the "break" came, it caused a terrific emotional storm, but once the storm subsided, the differences that followed caused far less disturbance than I ever thought possible.

Lessened anxiety; more flexibility; more positive outlook; interpersonal relations much improved; more effective—I especially notice this in my role as mother and as nurse; interests broadened; more direct and indirect satisfactions.

(1) Increase in emotional tone—being able to have "pure" strong feelings . . . now when I get angry I have a pure, strong feeling of anger. Expression of this anger brings relief. Anger makes me aggressive and makes me look for a solution and catharsis. Formerly, anger made me nervous, frightened, submissive, and eventually very depressed. (2) Insight—ability to become my own therapist

and to handle formerly frightening and inexplicable feelings or fantasies.

Increase in Self-Esteem

A fairly sizable percentage of the patients (38.5%) made specific reference to increments in self-esteem, self-respect, and self-acceptance. These changes, as well as those reported in the other categories, are congruent with benefits from all forms of psychotherapy—they are not unique to psychoanalysis or intensive, analytically oriented psychotherapy. The impression is frequently suggested by the data that improved self-acceptance is the *product* of deeper intrapsychic changes rather than a result in itself. As some of the quotations indicate, increases in self-esteem appear to be a function of a more realistic self-appraisal of personality assets and liabilities, a diminished need for perfectionism and narcissistic self-aggrandizement, and (probably through identification with the therapist) a more accurate assessment of and regard for the patient's own worth. These modifications therefore appear to be closely intertwined with those reported under the first rubric. A few direct quotations will illustrate the foregoing points:

> Self-esteem increased appreciably.
>
> As to internal changes, I have had more respect for my own ideas and abilities and I have been much more successful in "getting close" to people.
>
> I have learned to accept myself as I am, and I find I rather enjoy being me; the most drastic change, in other words, has been to learn that I can live an enjoyable life without any drastic changes. I do not mean that I now make no effort to improve myself; on the contrary, I think I am more intent on this than before. The difference is in my outlook and attitude more than anything else. As a result of the above, I find I have more self-confidence, and am willing to tackle things that I would not have attempted before, though I may not like it any better.
>
> I have more confidence in myself, in how I look, talk, etc.
>
> Finally, I do like myself more than I did in the beginning, and it is not as important to "will" myself to conform.
>
> The first important change is the fact that I am beginning to accept my own judgment with respect.
>
> I am at peace with myself. I no longer feel such a failure inside that I have to try to be 100% perfect on the outside and be all things to all men, so I can relax. I have discovered I'm more capable than I thought myself, for instance, I have learned to paint and in fact paid for my treatment with money I made myself. I might add that I'm no longer quite so popular since I can say "no" to my friends, but this doesn't bother me. I can pretty much take each day as it comes now and can derive much satisfaction from little things.
>
> Calmer; more accepting of myself and others.

Better Interpersonal Relations

Concomitant with significant intrapsychic changes, improvements in interpersonal relations were noted specifically by 28.2% of the respondents. There seems to be greater awareness of the needs, rights, and feelings of others, an ability to see them as people in their own right, greater freedom in approaching, responding, and interacting with them, and more give-and-take. In other words, the patients feel less of a compulsive need to fit others into a neurotic pattern. Also implied is a lessened need to dominate and to control (and conversely to be dominated and controlled). A few examples will serve:

> I have been much more successful in "getting close" to people.
>
> More understanding of others' problems and needs.
>
> Find that I am shy with people, especially girls, but not phobic as before.

One patient put it beautifully and succintly:

> Understanding of others—others formerly were mysterious people to me and I felt they knew more about me than I knew about myself. Now, I have an appreciation of the problems, fantasies, motivations of others after having learned a great deal about myself.

Improvement in Presenting Neurotic Symptoms

Over one half of the respondents commented on the disappearance or reduction of a variety of symptoms, both physical and emotional. Such comments typically were embodied in fuller accounts describing changes in general outlook and functioning. Symptoms no longer are seen as foreign bodies in the personality, but they have become better integrated in the total personality. Most noteworthy perhaps is the patients' changed attitude toward the symptoms themselves. They are no longer something passively endured but something to be actively dealt with and mastered. Even in instances where symptoms have not completely disappeared, the patient feels more comfortable about them. This greater openness to experience undoubtedly flows from deeper changes and reflects a reorganization of the personality. As one patient put it:

> The most helpful part to me I think is that I no longer try so desperately to hide hostility from myself—I usually try to accept it, let it go and move on. I have found that I have better family relationship, but still have a long way to go.

Implicit in this formulation is a greater willingness to make compromises where compromises have to be made, but also a determination to work toward viable solutions, even if perfection is not in prospect. A few additional quotations exemplify similar attitudes:

> The tight feeling in my chest over the heart has improved very much. I still have the effects of a feeling that my heart skips a beat at times when something has upset me. I still get tense when I can't get things done when they should be even though no fault of my own, but as soon as I realize what is happening I start to calm down.
>
> Am more able to deal with anxiety and tension in everyday life.
>
> Less distress about symptoms. Symptoms still exist, some to a less degree, but they do not loom as large as before therapy.

More Active in Working out Solutions

The changes classified under this heading appear to be byproducts of personality modifications which have already been commented upon, and need no further elaboration. They are often implied in other comments, and thus are probably underrepresented in the table. Nonetheless, they call attention to a rather basic change in personality functioning—from passivity to activity—which was heralded by Freud as one of the cardinal objectives of psychoanalytic therapy.

Current Adjustment and Current Problems

The question of improvements resulting from psychotherapy was approached somewhat more indirectly in several other ways. For example, the patient's current adjustment might give important clues concerning the resolution of his previous difficulties. Question 76 was addressed to this problem.

The response distribution indicates that 88.6% of the patients regarded themselves as getting along at least "fairly well." Of the total group, one half said that they were getting along "very well" or "extremely well." Nine percent described their adjustment as neither good nor poor, whereas 2.3% stated that they were getting along "very poorly." On the whole, these self-evaluations are extremely favorable.

76. On the whole how do you feel you are getting along now?

	n	%
Extremely well	6	13.6
Very well	16	36.4
Fairly well	17	38.6
Neither well nor poorly	4	9.1
Fairly poorly	0	0
Very poorly	1	2.3
Extremely poorly	0	0

This rosy picture, however, does not signify the patients experienced no further difficulties in living. We shall presently see that a variety of problems persist and presumably some new ones have been added. It may be assumed that a sample drawn from the general population might disclose results that are not too different from the foregoing. (For evidence on this point, see Gurin, Veroff, & Feld, 1960).

In Item 77 the respondent was asked to list the kinds of problems he feels exist now (unresolved and/or new ones), and Item 78 followed up on his statements by asking how adequately he was dealing with them. The responses to the open-ended items were classified independently of the responses to the precoded question. The results are presented in Table 5.

There are relatively few former patients (about 9%) who regard their problems in living as slight and who indicate a high degree of self-confidence in dealing with them. The possibility that such strong assertions may entail denial must be borne in mind. The group which admits to "mild" problems is probably expressing more realistic appraisals of their situation (approximately 37%). When we consider the responses to Item 78, we find that a fair proportion of the respondents, whose remaining problems are probably of moderate severity, believe they are dealing "fairly adequately" with their difficulties (63.7%). This leaves about 7–9% of the sample comprising people who consider themselves in a great deal of trouble which they find more than they can handle by themselves. These findings corroborate an observation which has already been made in a somewhat different context: the experience of psychotherapy, despite its great value to the patient, has left him with a fair number of unresolved intrapsychic and/or interpersonal problems. However—and this may be the greatest achievement of their therapeutic experience—they disclose appreciably greater inner strength and self-confidence in admitting, facing, and coping with their conflicts than they had at the beginning.

78. How adequately do you feel you are dealing with them?

	n	%
Very adequately	4	9.1
Fairly adequately	28	63.7
Neither adequately nor inadequately	6	13.6
Somewhat inadequately	3	6.8
Very inadequately	1	2.3
Blank	2	4.5

Need for Further Therapy

The patient's need for further therapy may be regarded as another indicator of his current emotional well-being, and the responses to this question closely parallel those discussed in the preceding section. In studying the responses to this question it must be kept in mind that most patients had been in psychotherapy for a fairly extensive period of time, and (with few exceptions) they undoubtedly had the experience that submitting one's thoughts, feelings, fears, and free associations to the scrutiny of a professional person can serve the purpose of clarifying one's emotions. Therefore, whenever such a person encounters difficulties in living he may wish for a therapeutic relationship in which his conflicts can be clarified and perhaps even resolved. Apart from a realistic need for further therapy, the answers to this question are undoubtedly permeated by feelings of this kind.

TABLE 5

SUMMARY OF UNRESOLVED AND/OR NEW PROBLEMS

($N = 43$)

Classification	*n*	%
No problems other than ordinary problems in living; coping adequately	4	9.3
Mild problems in living	16	37.2
Problems of moderate severity, but coping	19	44.2
Severe problems in living; great difficulty in coping	3	7.0
Unclassifiable	1	2.3
Total	43	100.0

Responses to Item 16 indicate that approximately one third (31.8%) of the patients indicated that they needed no further therapy; an additional 27.3% described their need as "slight," whereas 40.9% felt that they definitely needed more therapy. In our opinion, the response distribution for this item, while less favorable than those dealing more directly with satisfaction and benefits derived from therapy, represents a more realistic estimate of the patients' current emotional well-being.

16. How much in need of further therapy do you feel now?

	n	%
No need at all	14	31.8
Slight need	12	27.3
Could use more	8	18.2
Considerable need	10	22.7
Very great need	0	0

A slightly different slant on the same problem was obtained in Item 15, which was addressed to those patients who had had only one period of therapy (some 29% had had two or more therapists). The response distribution was substantially similar to that of Item 16, discussed in the foregoing. Patients who "never" or "very rarely" experienced the need for further therapy numbered 13 (29.6%); on the other hand, 38.6% felt a reasonably strong need for further treatment.

15. If you had only one period of therapy, did you ever feel a need for further therapy?

	n	%
Never	5	11.4
Very rarely	8	18.2
A number of times	13	29.5
Often	3	6.8
Very often	1	2.3
Blank	14	31.8

This question was followed by an open-ended item inquiring into the reasons for not seeking further help if the patient felt he needed it. Practical factors, such as distance from the treatment center and financial considerations were mentioned with some frequency. Some direct quotations of patients' responses will further illustrate their thoughts and feelings:

> The desire was not strong enough, and circumstances causing this desire were of short duration and infrequent.
>
> Felt I ought to work out own problems.
>
> Expense. There is a vast difference between a "need" and an "urgent need" for therapy, and I never really had the latter after the initial therapy was terminated.
>
> Though I am still considering myself in the one period of therapy, there have been times when I felt I'd like to have gone, but did not, chiefly because I did not want some of my family and friends to know I was going and therefore waited for easy convenient times to take the trip. The fact that I traveled about 30 miles each way was partly to blame—and because I just got *busy* and had trouble finding time.
>
> 1. Realized necessity of becoming independent. 2. Dislike of dependency on therapist. 3. Some anger at therapist's suggestion of termination.
>
> Too hard to express what I felt. Didn't think anyone would be really interested.
>
> I haven't felt the need and cannot imagine any problems arising that I could not work out fairly well. However, should such an occasion arise, I would not hesitate to seek professional help again —preferably in Chapel Hill. I found people there not only competent, but kind and helpful (I am not speaking of the therapist—rather, other personnel, secretaries, etc.). I think this warmth is very important from the patient's point of view and that this sort of moral support is very often overlooked by professionals in the field.
>
> 1. Financial, shortly after stopping. 2. Later I found it difficult to think in terms of returning to examine painful areas—it is a feeling of wanting to "let sleeping dogs lie." However, if I felt disturbed to the point of needing further professional help I would ask for further help.
>
> I did return to see my therapist on at least two occasions, one at his request when I re-entered school, the other upon leaving Chapel Hill. I seem to want to see him at times of change and growth. This spring I sometimes felt the need to see him when the demands of my new job seemed hard and caused a strain in outside relationships or when I felt depressed for several weeks at a time. Distance was one factor in not returning, also pride, and a question of motive. However, I hope to have several sessions with him during the second session of summer school and see if we can arrive at some answers. I imagine the act of answering the questionnaire will make patients "homesick" to a degree or perhaps prod them into raising their standards again for maximum development.

Correlates of Therapeutic Change

We shall now turn to a consideration of the relationships between Patient Cluster 2 (Amount of therapeutic change) and other attitude clusters derived from Questionnaire P. It will be noted from Table 6 that the amount of therapeutic change the patient attributes to himself shows a marked positive relationship with the degree to which the patient perceived the therapist as positively disposed toward him. In other words, the patient attributes to himself a great amount of favorable change when at the same time he rates in retrospect the therapist's attitude toward him during therapy as warm, respectful, natural and unstudied, attentive and the like (Cluster P_1). Coldness, distance, formality, extreme passivity, etc., by the same token, were rated as absent from the

TABLE 6

SELECTED CORRELATES OF AMOUNT OF THERAPEUTIC CHANGE (CLUSTER P_2)

Variable	*r*
Cluster P_1 : Therapist's warmth	$.53^{**}$
Cluster P_3 : Present adjustment	$.60^{**}$
Cluster P_4 : Amount of change apparent to others	$.52^{**}$
Cluster P_5 : Therapist's respect and interest	$.63^{**}$
Cluster P_6 : Uncertainty about therapist's feelings	$-.57^{**}$
Cluster P_8 : Use of technical terms	$-.39^{*}$
Cluster P_9 : Degree of disturbance before therapy	.26
Cluster P_{11} : Intensity of treatment	$-.09$
Age	$.40^{*}$
Occupation/education	$-.40^{*}$
Income	.17

* $p < .05$.
** $p < .01$.

therapist's behavior. The therapist's respect and interest (Cluster P_5) emerge as another important ingredient in ratings of favorable change, as does the absence of uncertainty about the therapist's attitude toward the patient during therapy (Cluster P_6).

A greater amount of change also coincides with more favorable ratings concerning the patient's current adjustment (exemplified by the absence of a need for further therapy, his ability to get along well now, and his feeling of adequacy in dealing with current problems—Cluster P_3). The therapeutic change, the patient feels, is markedly apparent to others, to whom he feels he would recommend psychotherapy as a beneficial treatment procedure (Cluster P_4).

The statistical relationships indicate that older patients tend to express themselves more favorably concerning the amount of therapeutic change ($r = .40$); furthermore, and somewhat unexpectedly, occupational and educational status of the patient tended to be negatively associated with the amount of change he attributed to himself ($r = -.40$). It is possible that patients higher on the occupational/educational scale were more critical in admitting to therapeutic change. It is also noteworthy that the more the therapist was perceived by the patient as using technical language, the *less* change the patient attributed to himself ($r = -.39$).

A warm, respectful, unstudied attitude on the part of the therapist, as perceived by the patient, seemed to be most propitious for favorable change, as seen by the patient in retrospect. This attitudinal factor appears to permeate the large majority of patient ratings, as indicated by the high degree of consistency with which the aforementioned correlates were repeated from cluster to cluster. Therapy in which a great deal of change has taken place, then, is surrounded, in the patient's view, by a large component of positive feelings. It will be instructive, by contrast, to comment on a cluster whose item composite deals with essentially negative attitudes toward the therapist.

TABLE 7

SELECTED CORRELATES OF UNCERTAINTY ABOUT THERAPIST'S FEELINGS (CLUSTER P_6)

Variable	r
Cluster P_1 : Therapist's warmth	$-.69^{**}$
Cluster P_2 : Amount of change	$-.57^{**}$
Cluster P_3 : Present adjustment	$-.46^{*}$
Cluster P_4 : Therapist's respect and interest	$-.61^{**}$
Cluster P_7 : Intensity of emotional experience	.23
Cluster P_8 : Use of technical terms	.26
Cluster P_{11} : Intensity of treatment	$.38^{*}$
Cluster P_{13} : General dissatisfaction	$.47^{**}$
Age	$-.43^{**}$
Item P_{44} : "Nothing the therapist said or did ever decreased my self-respect."	$-.36^{*}$
Item P_{28}: "A major emphasis in treatment was upon gestures, silences, shifts in my tone of voice, and bodily movements."	$.41^{**}$

* $p < .05$.
** $p < .01$.

Uncertainty about the Therapist's Feelings (Table 7)

Items included in this cluster (Cluster P_6) suggest indifference on the part of the therapist, lack of understanding of the patient's feelings, the patient's doubt about the therapist's "real" feelings, and two items (scored negatively) which, in extreme terms, asserted the opposite. What are the concomitants of this expressed attitude?

The greater the uncertainty about the therapist's feelings, the less warmth the patient perceived in the therapist's attitude (Cluster P_1). Similarly, a rather substantial negative relationship was demonstrated between feelings of uncertainty about the therapist's attitude and amount of change admitted by the patient (Cluster P_2). The patient's complaints about his current adjustment were associated with feelings of uncertainty about the therapist's attitude during treatment (Cluster P_3). The same was true of feelings of general dissatisfaction (Cluster P_{13}). Put otherwise, the worse the patient felt currently, the more he tended to blame the therapist (and the therapeutic relationship) for his state, and he did this by attributing impersonal attitudes toward the therapist, accusing him of never revealing his "true" feelings toward the patient, and showing general indifference. These results are substantially the inverse of the findings for Patient Cluster 2 (Amount of change) in that negative patient attitudes toward the therapist were generally associated with poorer outcome, less change, and greater patient dissatisfaction with the results.

Again, it was found that younger patients tended to experience more uncertainty. It is less surprising that patients who had been in intensive treatment (psychoanalysis) were more likely to be in doubt about the therapist's feelings toward them (Cluster P_{11}). These two findings appear to be interdependent because younger patients were more likely to have been in more intensive treatment (see below).

Emotional Intensity of Psychotherapy

It is a commonplace that the psychotherapeutic process lays bare a host of painful and unpleasant feelings in the patient. Following Menninger (1958) what the patient gains in psychotherapy and psychoanalysis is the product of his frustrations and disappointments. These, in a deep sense, are his therapeutic achievements. The more intensive forms of therapy, clearly, arouse deeper and more disturbing emotions which are undoubtedly a function of the deeper regression the patient experiences.

TABLE 8

Selected Correlates of Intensity of Therapy (Cluster P_{11})

Variable	r
Cluster P_1 : Therapist's warmth	−.23
Cluster P_4 : Amount of change apparent to others	−.29
Cluster P_6 : Uncertainty about therapist's feelings	.38*
Cluster P_7 : Emotional intensity of therapy	.35*
Cluster P_{12} : Importance of emotion in therapy	.37*
Item P_{44} : "Nothing the therapist said or did ever decreased my self-respect."	−.45**
Item P_{56} : "A major emphasis in treatment was upon gestures, silences, shifts in my tone of voice, and bodily movements."	.32
Item P_{57} : "I was almost never given any direct reassurances by the therapist."	.32
Sex (chi square)	5.47*
Occupation/education	.23
Income	.27
Age	−.20

* $p < .05$.
** $p < .01$.

The statistical relationships presented in Table 8 are highly consistent with this sketch of the nature of intensive therapy. For one thing, there was a significant positive relationship between the Intensity of treatment cluster (Cluster P_{11}), based on three reasonably objective items such as frequency of sessions per week, length of therapy (total number of hours), face-to-face position versus use of the couch, and the intensity of emotion experienced by the patient (Cluster P_7). Furthermore, there was a tendency for patients who had undergone intensive therapy to report greater uncertainty about the therapist's feelings and attitudes. In other words, they experienced him as relatively anonymous and as a blank screen (Cluster P_6). As already noted, younger patients were more likely to receive intensive therapy (probably in line with the criteria of suitability outlined by Freud many years ago). Furthermore, it is important to note that patients who were seen in intensive therapy tended to *disagree* with Item 30 which stated:

> Nothing the therapist said or did ever decreased my self-respect.

We may safely infer that the latter statistical association refers to the therapist's occasional penetration of the patient's narcissistic defenses. Finally, such patients were more deeply impressed with the *emotional* experience of psychotherapy as opposed to *intellectual* understanding.

Another concomitant of intensive psychotherapy, as seen by the patient, was the therapist's emphasis on gestures, silences, shifts in tone of voice, and bodily movements (Item 56). This (psychoanalytic) stress on nonverbal communications, which presumably were at least occasionally the focal point of the therapist's interpretations, tended to be seen as incompatible with feelings of warmth on the part of the therapist. The absence of reassurance was also noted by patients in intensive therapy (Item 57).

The question may be raised as to whether patients who had been seen in intensive psychotherapy experienced greater change as a result of treatment than other patients. This hypothesis seems reasonable in the light of the supposition that greater (more intensive) therapeutic effort in terms of length of treatment, frequency of visits, etc., might be commensurate with greater therapeutic change. Our data provide little support for this view. The pertinent statistical relationships hovered in the vicinity of zero and indeed tended to be slightly negative. In other words, we found no evidence that therapists considered such patients as having been more successful cases, nor do the patients themselves credit themselves with greater change. Not surprisingly, patients who had been seen in intensive therapy were considered to have greater capacity for insight by their therapists ($r = .47$), and they tended to belong to the higher socioeconomic classes ($r = .23$, not significant). There also was a slight tendency for such patients to be younger and male ($\chi^2 = 5.47, p < .05$).

Attitudes toward the Therapist

At the beginning of therapy patients' attitudes toward the therapist covered a wide range. They reported being variously reserved, suspicious, hostile, resentful, indifferent, eager to cooperate, curious, receptive, anxious, and dependent. At the time of termination the patients' attitudes (at least in retrospect) had markedly changed, but in various ways. Some patients apparently attempted to arrive at a reasonable assessment of the therapist as another human being who, in his therapeutic role, had taught them better ways of looking at themselves and their problems, for whom they had gained a good deal of respect and fondness, to whom they were grateful for his dedication and steadfastness, whose impartiality they admired, and with many of whose positive attributes of maturity they had formed a permanent identification. There were other former patients whose attitudes might be best described as hero worship or idolization; and there were a few patients whose negative (or ambivalent) attitudes seemed to have become more intense during therapy. A few salient quotations will illustrate the range:

> The attitude was much more realistic. I no longer romanticized him as much, but I liked him more. I felt more on an equal plane with him. At the same time I was aware of more hostility and even had one big tantrum in which I kicked the floor. Although I was behaving like a two-year-old, we were both quite proud of this. I did feel rejected at the end of therapy though I knew that was unrealistic.
>
> I found him to be warm, understanding and a very "down to earth" person. I considered him a friend and ally.
>
> Confidence and trust and respect for him, and gratitude.
>
> I had a great deal of respect for him and I was grateful for his help, but no longer felt the exaggerated regard that I had during the early part of therapy.
>
> Great liking for therapist. Feel like a life-long bond with him. Admire him for his integrity, genuine desire to help people, considerateness, desire to search for truth, his sense of humor. Great respect for his insight, experience, intelligence. Complete confidence in him and no feelings of hesitation about verbalizing or expressing any feelings or thoughts.
>
> Near the end of therapy it was almost unbearable to think of the relationship ending—I loved and respected the therapist and did not in one sense want to say good-bye. I left with sadness but a heart full of thankfulness for all he had helped me to find, and I wanted so much to be able to give something in return for the healing which had come to me.
>
> None better. Fortune was entirely on my side when I found him. Never thought a mere human could be so wonderful.
>
> More hostile.
>
> Disappointed.
>
> I think there was little change.
>
> I loathed him personally; respected his ability.

Patient's Perception of the Therapist's Attitudes during Therapy

In Item P_{84} we attempted to elicit descriptions of the therapist's attitude as the patient experienced it during therapy. The answers occasioned no surprise because they seemed to coincide with the way in which the therapist's professional attitude is usually depicted. The following characterizations convey the composite: warm, sympathetic, anxious to help, respectful, attentive, interested, observant, detached, helpful, nonjudgmental, firm, sincere, encouraging, kind, stimulating, patient, honest, understanding, considerate, objective, accepting, cooperative, permissive, noninvolved, skillful, determined, sense of humor, concerned, unhurried, relaxed, consistent.

These adjectives appear to be a blend of personal attributes (friendly, kind, etc.) expressed in the framework of professional objectivity and detachment.

Some patients took exception to these attitudes or detected admixtures which they interpreted in negative terms. Some examples follow: passive, dull, vague, grave, somewhat overintellectual, somewhat bored, a sense of pathos transcending empathy, unattentive, bemused, professional, superficial pleasantries but still removed, strictly adhering to schedule, rather immature attempts to maintain "proper" facade of objective, impersonal arm's length coolness; I believe he tried to appear neutral toward me but to me it seemed that he was rather indifferent, I was not always sure but in general I felt he was fairly sincere in his desire to help.

Nature of Therapeutic Change as Seen by the Patient

Item P_{85} was aimed at assessing the patients' perception of the factors responsible for therapeutic change. It read:

> What specific aspects of the therapist's activities do you feel accounted for the change as you experienced it?

The responses to this item, like the responses to many items in the questionnaire, cannot be taken at face value; that is, one would not assume a one-to-one relationship between the patients' descriptions of the nature of the therapeutic influence and the "true" factors to which therapeutic change is attributable. At the same time, these reports no doubt reflect *some* important aspects of the experience which the respondent found helpful.

From the patients' responses, it appears that a useful distinction may be made between *technical interventions* (interpretations, clarification, etc.) and *personality characteristics* of the therapist which enabled the patient to view himself and his emotional difficulties in a different light. The latter group includes the therapist's human qualities (for example, interest, respect, trustworthiness, patience, acceptance, firmness, and reliability) as well as a host of attitudes toward the patient and his problems. It is clear that through his attitudes and through his personality the therapist educates the patient and provides him with a model for identification. Although these accounts typically came from our more intelligent and more highly educated persons, they are the formulations of patients—not professionally trained people. Thus, they are relatively uncontaminated by the technical literature although a certain amount of "indoctrination" by the therapist has undoubtedly occurred. The following quotations exemplify these assertions:

> Gave me a clear picture of my emotional problem. Gave me personal confidence.
>
> It is hard to be specific, but I think two things stand out: (1) He pointed out things I did or said, from time to time, which frequently put them in a different light; (2) The rest of the time his attitude seemed to say: "It's your move," which brought pressure on me to take the initiative.
>
> Getting me to talk about and face honestly feelings, fears, concerns, listening to me without approving or disapproving. Helping me to interpret things which bothered me but which I did not understand—bodily gestures, dreams, feelings of anger. His quiet unhurried manner.
>
> His seeming understanding. His apparent sympathy. His refusal to become subordinated by my personality. His way of making me feel better before I left each session—some new light cast, etc.
>
> His listening, his being serious or laughing at the right times. A certain amount of the didactic process when it was necessary. A refusal to think I was irresponsible or "sick." A belief that the achievement of freedom re life and its outcome is possible even for the troubled, in others words, his philosophy of life.
>
> His lack of surprise or shock. His apparent understanding and respect for me. His ignoring or

belittling some of the things that worried me so that I realized they weren't so important as I'd thought. His actual scientific statements of fact on some subjects....

His persistence and my feelings he would stick it out for however long it took to accomplish what was needed.

Never got shocked at anything I said. I was able to see that some of my innermost feelings of guilt and "badness" were not really bad but were natural reactions to early childhood situations of a traumatic nature. His encouragement to explore my phantasies, get behind memory screens, etc. He helped "trigger" latent feelings by key questions. Pointed out where I was getting off the track when I switched subjects. His summaries of therapeutic hours and putting parts and trends together, but never getting too far ahead of me or at least revealing this.

Her genuine human qualities; her deep desire to help me in spite of myself.

Clarification of feelings—pointing out similarities of feelings in various circumstances.

I am somewhat undecided about this; however, I do feel that his delving into my childhood experiences helped alleviate a number of unresolved guilt feelings; on the whole, however, I am not certain what aspect of his activities resulted in change.

Employing the distinction mentioned above, these comments are largely in terms of the therapist's *personality characteristics* rather than in terms of *technical intervention*, although references to technical maneuvers are sometimes implied. Clearly, what most impressed our respondents was the therapist's personality and attitudes.

Patients' Identification with Therapists' Values

A follow-up question to Item 85 inquired specifically about the degree to which the patient felt he had "taken over" the therapist's attitudes, values, or opinions. Seventeen patients answered the question in the affirmative, whereas 22 denied it and omitted the item. Further clarification is provided through the following direct quotations which are considered typical for the sample:

No. He had no opinions.

Unable to recall anything definite.

I don't know because he did not reveal his own attitudes, values, and opinions.

In most cases, I did not feel that I knew my therapists' attitudes, values, opinions. I had throughout and now a general feeling that what he said was not so much what he thought, but more to bait or challenge me, or thrust me into a new reality for which I was not prepared, and which I in some cases would never need to be aware of.

Yes, commonsense attitude toward some basic emotional problems.

Tolerance.

To be realistic, understanding; not to be emotional.

Sense of humor; belief in change; reverence for life (not just lip service); a greater understanding of others—the "why" of them; increased ability to reason things out; an appreciation and respect for emotions.

Yes! To learn the extent of one's abilities and to accept the extent of their limits as the best one can do.

Not to keep feelings bottled up inside; not to expect or demand perfection in everything; to fail in one's efforts sometimes is no disgrace.

Are Therapeutic Changes Apparent to Others?

No clear-cut patterns were discernible with respect to the amount of change that the patient thought might be apparent to others, although there seemed to be an overall tendency for changes to be reported as most apparent to intimate associates (husband, wife) and least apparent to co-workers and acquaintances. As is shown in Table 9, the extent to which the changes resulting from therapy were described as apparent to an outsider was associated with overall amount of change experience by the patient himself. The degree to which these changes seemed apparent to others was negatively correlated with the patient's educational level and his capacity for insight as rated by the therapist. Thus, it appears that the less educated patients and those whose capacity for insight (and hence intensive therapy) was judged low were the ones who reported the most dramatic changes as far as improvement in interpersonal relations was concerned.

Socioenvironmental variables

While most of the statistical relationships between socioenvironmental variables and attitude clusters were inconclusive, certain statistically significant associations emerged, which are summarized in Table 10.

TABLE 9

SELECTED CORRELATES OF IMPROVEMENT IN INTERPERSONAL RELATIONS

Variable	Item P_{75}: Amount of change apparent to: Husband or wife	Close friends	Co-workers and acquaintances
Item T_{5a} : Capacity for insight—before	−.41**	−.35*	−.35*
Item P_5 : Education	−.50**	−.45**	−.49**
Item P_{73} : Amount of change as result of therapy (as rated by patient)	.76**	.51**	.59**
Item P_{79} : Extent of symptom relief (as rated by patient)	.54**	.33*	.38*

* $p < .05$.
** $p < .01$.

The most important of these are the following:

1. Age. Older patients tended to attribute greater therapeutic change to themselves ($r = .40$); they felt the changes resulting from psychotherapy were more important to others ($r = .31$); they tended to feel *less* uncertain about the therapist's feelings toward them ($r = -.43$); and they were less likely to see therapy as an intensely emotional experience ($r = -.37$).

2. Sex. Women were less likely than men to be seen in intensive therapy ($\chi^2 = 5.47$); women were also seen as feeling more warmth for their therapists.

3. Family income. There was a positive relationship between size of family income and extent of remaining disturbance, as rated by the therapist ($r = .34$), signifying that persons with higher income tended to remain more disturbed. As already noted, there was a trend ($r = .27$, approaching statistical significance) for the more affluent patients to be seen in

TABLE 10

SELECTED CORRELATES OF SOCIOENVIRONMENTAL VARIABLES

Variable	Present age (r)	Sex (χ^2)	Family income (r)	Occupation/Education (r)
Cluster T_1 : Overall success	.11	1.13	.03	.11
Cluster T_2 : Remaining disturbance	−.09	.02	.34*	−.08
Cluster T_4 : Capacity for intensive therapy	−.23	2.83	.16	.41**
Cluster T_5 : Adjustment before entering therapy	.06	1.90	−.14	.31*
Item T_{1a} : Defensiveness before entering therapy	.23	2.93	.26	−.32
Item T_{16} : Patient warmth	.24	4.24*	.01	−.01
Cluster P_1 : Warmth of therapist	.26	.24	−.22	−.15
Cluster P_2 : Amount of change	.40*	2.06	.17	−.40*
Cluster P_3 : Present adjustment	.28	.14	.06	−.23
Cluster P_4 : Amount of change apparent to others	.31*	.02	−.17	−.54**
Cluster P_5 : Therapist's respect and interest	.26	.05	−.06	.16
Cluster P_6 : Uncertainty about therapist's feelings	−.43**	2.21	−.03	.05
Cluster P_7 : Intensity of emotional experience	−.37*	2.47	.08	−.01
Cluster P_{11} : Intensity of treatment	−.20	5.47*	.27	.23

* $p < .05$.
** $p < .01$.

TABLE 11

SELECTED CORRELATES OF THERAPISTS' ASSESSMENTS

Variable	Cluster T_1: Overall success	Cluster T_2: Remaining disturbance	Cluster T_3: Warmth of relationship	Item T_8: Motivation before therapy	Item T_{18}: Quality of working relationship
Cluster T_2: Remaining disturbance	−.66**	—	−.25	−.16	−.66**
Cluster T_3: Warmth of relationship	.54**	−.25	—	.24	.56**
Cluster T_4: Capacity for intensive therapy	.26	−.16	.44**	.22	.29
Cluster T_5: Patient's adjustment before therapy	−.09	−.30	.08	−.40**	.06
Item T_{1a}: Defensiveness before therapy	−.26	.45**	−.10	−.13	−.30
Item T_8: Motivation before therapy	.35*	−.16	.24	—	.29
Item T_{16}: Patient's warmth toward therapist	.75**	−.47**	.65**	.28	.67**
Item T_{18}: Quality of working relationship	.71**	−.66**	.56**	.29	—

* $p < .05$.
** $p < .01$.

intensive therapy. Such persons also seemed to be more inclined to recommend therapy to their friends ($r = .31$).

4. Occupation/education. Patients higher on the occupation/education scale were seen by their therapists as having greater capacity for intensive therapy ($r = .41$), but there was only a slight trend suggesting that they were in fact more likely to be accepted for intensive therapy ($r = .23$, not statistically significant). On the other hand, better educated patients had a better adjustment "before" therapy ($r = .31$), and they were viewed by their therapists as less defensive "before" therapy ($r = -.32$). As already noted, the higher the patient placed on this scale, the *less* he was prone to credit himself with marked therapeutic change ($r = -.40$). There was a rather substantial negative association ($r = -.54$) between the patient's position on the occupation/education scale and the amount of therapeutic change he believed was apparent to others.

B. *Therapist's View*

How do the participating therapists regard the patients who responded to the questionnaire? What are the concomitants of the therapists' retrospective ratings? To obtain a preliminary answer to this question we shall examine the interrelationships between clusters of items based solely on therapists' ratings as well as some individual items (see Table 11).

As might be expected, therapists consider successful treatment outcome (Cluster T_1) incompatible with remaining disturbance (Cluster T_2, $r = -.66$). As already noted, a good, warm, working relationship was seen as a major ingredient in successful outcome (success versus quality of the working relationship = .71, and success versus patient's warmth = .75).

The patients' capacity for intensive therapy (Cluster T_4) bore a small positive relationship to overall success ($r = .26$, not statistically significant); but, contrary to the patients' self-ratings (see above), it was not seen as contraindicating a warm relationship ($r = .44$). This rating of "capacity" of course was not necessarily a guarantee that the patient was in fact treated by intensive psychotherapy, but therapists derived greater satisfaction from patients they saw in intensive therapy. (Item T_{22}, "Enjoy working with this kind of patient" versus Item P_{14}, "Face-to-face position versus couch," $r = .50$).

Patients who in the therapist's view

were better adjusted "before" therapy (Cluster T_5) tended to be seen as less highly motivated for treatment (Item T_8, $r = -.40$). Since the Adjustment cluster included items pertaining to anxiety, ego strength, and degree of disturbance, the foregoing finding suggests that sicker patients are more highly motivated for therapy. Since motivation for therapy was seen as an ingredient in overall success ($r = .35$), one would expect that better adjusted patients, being less well motivated, would have poorer outcomes. However, no demonstrable relationship exists between Adjustment before therapy (T_5) and Overall success (T_1 ; $r = -.09$). This leaves the question unanswered, as far as our data are concerned.

Patients who were considered highly "defensive" before therapy (Item T_{1a}) tended to be viewed as more disturbed upon termination ($r = .45$). Also, defensive patients were rated as showing less warmth toward the therapist ($r = -.31$, not shown in Table 11).

The relationships between therapists' ratings and socioenvironmental variables were for the most part inconclusive. Among the exceptions, the following deserve mention (see also Table 10): There was a tendency for patients with higher family incomes to be rated as remaining *more* disturbed following therapy ($r = .34$); such patients also tended to be seen as more defensive ($r = .26$, not significant). There was a suggestion (barely below the minimum level of statistical significance) that older patients established a better working relationship with their therapists ($r = .29$). Female patients were rated by their therapists as showing more warmth toward them ($\chi^2 = 4.24$). Patients higher on the occupational-educational scale were shown to have greater capacity for intensive therapy ($r = .41$). The latter group of patients was seen as better adjusted before therapy ($r = .31$).

To summarize, the foregoing results suggest that there are two major ingredients to successful outcome (from the therapist's standpoint). The first of these is the quality of the relationship the patient establishes with his therapist, and the second is the patient's motivation at the beginning of therapy. Since the therapeutic relationship is the therapist's main source of information about the patient, and thus reflects most sensitively the patient's current status, it is to be expected that the therapist's rating of the outcome of therapy will in large part be determined by the quality of the relationship.

The other major ingredient of successful therapy—the patient's motivation before therapy—seems to be based on a global rating of the patient's status at the beginning of therapy. Undoubtedly, the therapist's rating of the patient's motivation is influenced by the patient's age, education, income, degree of disturbance, and many other factors. However, none of these measures, taken alone, is predictive of success. Apparently, the patient's assets and circumstances at the beginning of therapy interact in a complex way, so that in each case their relative importance may change.

C. *Patients' and Therapists' Views of Each Other*

Thus far we have investigated patients' retrospective accounts of their therapy experience as well as the interrelationships among their feelings and attitudes toward the therapist and the therapeutic experience. We have also reported data on certain evaluations of these patients by their therapists, explored the interrelationships among these ratings, and studied the evaluations in relation to selected socioenvironmental variables of the patients. We shall next turn to an exploration of the degree of agreement between the two sets of data. These analyses are the core of the present investigation in that they hopefully provide evidence regarding the degree of consensus between the participants in individual psychotherapy. How do therapist and patient feel about each other in retrospect? Is there any agreement between therapists and patients as to the benefits derived from psychotherapy? In a sense the data which answer these questions provide a form of validation of the

TABLE 12

SELECTED RELATIONSHIPS BETWEEN PATIENTS' AND THERAPISTS' ASSESSMENTS (I)

Patients' assessments and occupation/education	Therapists' assessments:			
	Cluster T_1: Overall success	Cluster T_2: Remaining disturbance	Cluster T_4: Adjustment before therapy	Item T_{18}: Quality of working relationship
Cluster P_1: Therapist's warmth, attentiveness, integrity, patient's trust	.44**	−.46**	.11	.48**
Cluster P_2: Amount of change	.36*	−.23	−.13	.39*
Cluster P_3: Present adjustment	.30	−.45*	.15	.16
Cluster P_5: Therapist's respect and interest	.59**	−.59**	.28	.69**
Cluster P_6: Uncertainty about therapist's feelings, lack of acceptance	−.43**	.38*	−.22	−.53**
Cluster P_9: Degree of disturbance before therapy	.53**	−.41**	−.09	.31*
Occupation/education	.11	−.08	.31*	.02

$^{*}p$ .05.
$^{**}p < .01$.

assertions made independently by the patient and therapist. Of course, they do not "prove" in a decisive way that the relationship had in fact the characteristics which the participants attributed to it, but they provide (in the terminology of Harry Stack Sullivan) "consensual validation" of the therapeutic experience.

Therapeutic Outcome

The therapist's rating of overall success may be accepted as the single most informative rating of therapeutic outcome. Although the shortcomings of such ratings are recognized, they are the best we have. Table 12 discloses that overall success as rated by therapists shows several substantial correlations with parallel patient assessments, most notably with Patient Cluster 5 (Therapist's respect and interest; $r = .59$). At a somewhat lower level ($r = .36$), therapists and patients agree concerning outcome as rated by the therapist and the amount of change (Patient Cluster 2) as rated by the patient. The therapist's warmth, attentiveness, integrity, and the patient's trust in these qualities (Patient Cluster 1) likewise were found to correlate significantly with ratings of overall success ($r = .44$). In contrast, uncertainty about the therapist's feelings, lack of acceptance, and deficient interest on the part of the therapist (Patient Cluster 6) correlated negatively with overall success as rated by the therapist ($r = -.43$).

Finally the degree of disturbance (Cluster P_9) which brought the patient to psychotherapy was markedly correlated with the therapist's rating of outcome ($r = .53$). In other words, patients who regarded themselves as seriously disturbed were more likely to be rated by the therapist as having been a therapeutic success.

Therapist Cluster 2 (Remaining disturbance) represents essentially the obverse of overall success ($r = -.66$) and yields findings largely reciprocal to the ones just reported. Thus, patients whose remaining disturbance was rated high by the therapist were said to have had a poorer therapeutic relationship ($r = -.66$) and to have been more defensive before therapy ($r = .45$). They were also described as not feeling warmly toward the therapist ($r = -.47$), and the therapist tended to see the tenor of the relationship as unpleasant ($r = -.54$).

Turning to the patient's point of view, these same individuals reported that they did not feel assured of the therapist's respect and interest ($r = -.59$); that they saw the therapist as deficient in warmth, attentiveness, and integrity ($r = -.46$); that they were uncertain about the therapist's feelings; and that they felt a lack of interest and acceptance on the therapist's part ($r = .38$). They also saw their present adjustment as poor and felt in need of further therapy ($r = -.45$). Again we see that greater disturbance before therapy as seen by the patient is correlated with less remaining disturbance as seen by the therapist ($r = -.41$). These findings again point to the operation of a general "warmth" or "positive attitudes" factor in both the therapist and the patient. However, the findings presented in Table 12 document not only a consensus between therapist and patient in terms of the quality of their relationship and their attitudes toward each other, but they also indicate marked convergence on clinical assessments which seems to go far beyond mutual liking. Despite their different vantage points, patients and therapists show reasonable agreement on the patient's psychological status after therapy.

While the evidence supports the proposition that greater disturbance upon entering therapy foreshadows better therapeutic results (perhaps because more improvement is possible for a sicker patient) it should be pointed out that there was no consensus between patient and therapist with respect to the degree of the patient's disturbance ($r = .09$) at the beginning of therapy. This may be explained in terms of differing conceptions regarding "disturbance." Cluster P_9, as will be noted from the item composite, is not a pure measure of disturbance; in addition, it seems to reflect the patient's eagerness to enter therapy and thus is probably an index of his motivation for treatment as well. Patients with better therapeutic outcomes, then, could be regarded as in some sense more highly motivated for treatment than patients who had poorer outcomes. On the other hand, the therapist undoubtedly compared a given patient with other patients he had seen; from *his* point of view, a neurotic patient, even if he is suffering from fairly intense anxiety, but is highly motivated for treatment, may be rated as far less seriously disturbed than a chronic schizophrenic patient in a state hospital.

Quality of the Working Relationship

There was a substantial association ($r = .69$) between the therapist's and the patient's assessment of the quality of the working relationship in therapy (see Table 12). The therapist's respect, his genuine interest in helping, and his communication of the attitude that the patient was worthwhile as a person seem to be key ingredients in this statistical finding. It is also noteworthy that this represents a shared attitude; stated otherwise, therapist and patient were not in doubt that a good working relationship existed when it did, nor were they deceived when it was deficient. We consider it important to point out that while the items in Patient Cluster 1 (Warmth) singly and in a composite correlated significantly with overall success and quality of the working relationship as rated by the therapist, they did so consistently at a lower level of magnitude than did Cluster 5 which highlighted the patient's feelings of being respected and accepted as a worthwhile person. If this represents a "true" (replicable) difference, it may indicate that the therapist's respect is a more important ingredient in success than the characteristics subsumed under the warmth cluster. Incidentally, it is to be noted that therapists were reasonably successful in gauging the degree of warmth the patients felt toward them ($r = .54$).

Whereas the therapist's assessment of the patient's motivation for therapy correlated moderately with his ratings of overall success ($r = .35$), there was little evidence that the therapist's rating of the patient's motivation (see Table 13) was a good predictor of the kinds of attitudes which they developed toward each other or the amount of benefit the *patient* felt he derived from therapy. A high level of motivation on the patient's part at the beginning of therapy, while predisposing toward a favorable outcome, does not in and of itself guarantee that he will form a productive relationship with his therapist.

TABLE 13

SELECTED RELATIONSHIPS BETWEEN PATIENTS' AND THERAPISTS' ASSESSMENTS (II)

Patients' assessments	Therapists' assessments: Item T_8: Motivation before therapy	Item T_{16}: Patient's warmth	Item T_{23}: Pleasantness/unpleasantness
Cluster P_1: Therapist's warmth	.30	.46**	.42*
Cluster P_2: Amount of change	.18	.41*	.38*
Cluster P_5: Therapist's respect and interest	.22	.54**	.55**
Cluster P_6: Uncertainty about therpist's feelings	−.24	−.50**	−.33*

* $p < .05$.
** $p < .01$.

The therapist's ratings of overall pleasantness/unpleasantness correlated positively with the amount of change the patient experienced ($r = .38$), the therapist's warmth as seen by the patient ($r = .42$), his respect and interest ($r = .55$), and negatively with the patient's feelings of uncertainty about the therapist's feelings ($r = -.33$). It is concluded that therapists derived greater enjoyment from working with patients who in turn felt favorably toward the therapist and who at the same time felt improved.

TABLE 14

SELECTED CORRELATES OF FORM OF THERAPY

Item	r
Item P_{38}: The therapist tended to be rather stiff and formal.	.35*
Item P_{41}: I feel he was extremely passive.	.48**
Item P_{44}: Nothing the therapist said or did ever decreased my self-respect.	−.33*
Item P_{50}: The tone of his statements tended to be rather neutral.	.39*
Item P_{56}: A major emphasis in treatment was upon gestures, silences, shifts in my tone of voice, and bodily movements.	.35*
Item P_{61}: I never had the slightest doubt about the therapist's interest in helping me.	−.34*
Item P_{62}: I was often uncertain about the therapist's real feelings toward me.	.35*
Item P_{63}: The therapist's manner of speaking seemed rather formal.	.44**

Note.—Item T_{20}: Rating Scale completed by therapist assessing the intensity of treatment on a linear scale ranging from 1 (supportive) to 5 (intensive-analytical).
* $p < .05$.
** $p < .01$.

Therapist's Attitude as Experienced by the Patient

From the results thus far reported concerning patients' perceptions, it appears that there is a continuum of therapist behaviors, ranging from the extremes of coldness, remoteness, and distance to warmth, nearness, and humanness. Our results show that the patient's placement of his therapist on this continuum is related to the intensity or "depth" of the treatment the therapist employed (see Table 14). If the therapist was employing realtively intensive, analytically oriented treatment procedures, the patient perceived him as stiff and formal, passive, making statements which were neutral in tone, and as tending to decrease his (the patient's) self-respect. There was an emphasis on gestures, silences, and other nonverbal communications. The patient admitted doubts about the therapist's interest in helping him, and he was uncertain about the therapist's real feelings. These results are quite consistent with existing knowledge of different forms of therapy. Certainly the role behavior of the therapist in once-a-week therapy involving face-to-face contact with the patient is very different from that in psychoanalysis, and one may expect that

TABLE 15

INTERRELATIONSHIPS AMONG SELECTED THERAPIST ASSESSMENTS

Variable	Cluster T_5: Adjustment before therapy	Item T_{1a}: Defensiveness before therapy	Item T_8: Motivation before therapy
Cluster T_4 : Capacity for intensive therapy	.00	−.15	.22
Cluster T_5 : Adjustment before therapy	—	−.11	−.40**
Item T_{1a} : Defensiveness before therapy		—	−.13

** $p < .01$.

the therapist is experienced very differently by patients in these situations.

Since we have noted previously that the outcome of therapy is rated lower when the patient perceives the therapist as cold, distant, lacking in interest, etc., the findings discussed in the foregoing paragraph might suggest that patients seen in more intensive therapy would have had poorer outcomes. Our data, however, do not support this implication: intensity of treatment (Cluster P_{11}) and outcome of therapy (Cluster T_1) are essentially uncorrelated. The facts of the matter are that the therapist's warmth is related both to the intensity of the treatment and to the outcome of therapy (positively), but there is no demonstrable statistical relationship between intensity of therapy and outcome. A conservative interpretation of these results would suggest that the existence of a generally positive relationship between therapist and patient is the best single predictor of therapeutic success. Intensity of therapy, taken alone, is not a good predictor of success. Rather it may be regarded as one of a number of variables which, taken together, determine the character of the therapeutic relationship and ultimately its outcome. Our findings strongly corroborate the widely held belief that the character of the therapeutic relationship is the focal point through which all other variables operate.

"Clinical" Ratings in Relation to Patient Evaluations

Although many therapist assessments, whether attitudinal or clinical, tended to be highly intercorrelated, there were exceptions. Table 15 shows the interrelationship among items or clusters which appeared to possess a certain uniqueness. What are the relationships between these relatively objective clinical assessments and patient evaluations (see Table 16)?

It is not surprising that patients whose capacity for intensive therapy was rated

TABLE 16

SELECTED CORRELATES BETWEEN THERAPISTS' CLINICAL ASSESSMENTS AND PATIENTS' ASSESSMENTS

Variable	Cluster T_4: Capacity for intensive therapy	Cluster T_5: Adjustment before therapy	Item T_{1a}: Defensiveness before therapy	Item T_8: Motivation before therapy
Cluster P_1 : Therapist's warmth	−.23	.11	−.43**	.30
Cluster P_4 : Amount of change apparent to others	−.41*	−.19	.04	−.01
Cluster P_5 : Therapist's respect and interest	.18	.28	−.41*	.22
Cluster P_7 : Intensity of emotional experience	.48**	−.49**	.10	.28
Cluster P_{11} : Intensity of treatment	.47**	−.14	.05	.24

* $p < .05$.
** $p < .01$.

high by the therapist (Cluster T_4) were in fact accepted for this form of treatment (Cluster P_{11}). Moreover, patients whose capacity for treatment was rated high tended to describe their therapy experience as intensely emotional, painful, etc. (Cluster P_7, $r = .48$). Another concomitant of the capacity for intensive therapy emerged in the negative statistical relationship to Patient Cluster 4 (Amount of change apparent to others, $r = -.41$). This suggests that the changes experienced by patients in intensive therapy were predominantly intrapsychic and had less to do with their adjustment to reality on a macroscopic level.

The healthier the patient's adjustment before therapy (as judged by the therapist, Cluster T_5), the less he was likely to experience therapy as an intense emotional experience ($r = -.49$). This finding probably indicates that healthier patients were less likely to be accepted for intensive therapy, and it coincides with the finding that such patients were considered less highly motivated for therapy.

On the other hand, the more defensive the patient was judged "before" therapy (Item T_{1a}), the less likely he was to experience the therapist as warm (Cluster P_1, $r = -.43$). Similarly, the patient did not experience the therapist as helpful, interested, and respectful (Cluster P_5, $r = -.41$). Such patients evidently formed poorer relationships with the therapist.

As already noted, none of the correlations between the therapists' ratings of the patients' motivation and patients' self-ratings reached statistical significance, although Patient Cluster 1 (Warmth) and Patient Cluster 7 (Emotional intensity of therapy) came near. What is apparently operating here is a mutual dislike on the part of the patient and therapist: the patient presents himself in a defensive and unattractive way to the therapist, and the therapist responds by being disinterested and possibly by thinking of the patient in derogatory "clinical" terms—i.e., that he is a "poor therapeutic risk" and the like.

DISCUSSION

The present study was undertaken for the purpose of providing preliminary data concerning the therapeutic experience as the patient viewed it in retrospect and as the therapist evaluated it from the same perspective. It is amply apparent that it does not answer the question of *how* a given therapeutic result was achieved; nor of course does it lend itself to conclusions concerning the effectiveness of the psychotherapeutic methods employed by the various therapists. Its aim was primarily to describe and to explore.

Unfortunately, in this area of research there is a wide discrepancy between the researcher's aspirations and the practicalities of the situation. In principle, it should be possible to collect data over the course of therapy from therapists and patients. In practice it turns out to be an exceedingly difficult task. The design of the present investigation was frankly a compromise between what we should have liked to do and what it was possible to do. Even so, collecting the data reported in this paper proved to be an arduous task. Such difficulties are not particularly noteworthy in and of themselves because technical problems arise in all areas of research. Moreover, this is not the place to examine the complex set of factors impeding the collection of comprehensive data on psychotherapy. It seems important to record, however, that good data in this area are hard to obtain.

To illustrate, it must at least be mentioned that an appreciable proportion of the patients whose cooperation was invited failed to return completed questionnaires. This is not surprising and, indeed, was to be expected, given the types of topics covered and particularly that it was considered unfeasible to send follow-up letters. Some data have been presented to indicate that the returnee sample, probably as a result of these factors, may have been somewhat biased in the sense of giving a somewhat stronger representation of patients whose treatment was successful.

It is also important to note that the data do not allow one to follow the step-by-step progress in the treatment of these patients, nor do they provide firm and reliable anchoring points for objective evaluations of change. To meet this latter requirement one would have to obtain comprehensive diagnostic assessments at the beginning and at the end of therapy as well as after a follow-up period. In addition, one would have to gather detailed information about the character of the patient-therapist interaction over the course of treatment. Even under those circumstances, however, it would be difficult to attribute specific changes to specific events in therapy. Finally we must mention for a final time that this study suffers from the obvious drawbacks of retrospective ratings.

Given the aims of this investigation, its defects are somewhat less damaging than they might otherwise be. Our primary purpose has been to shed some light on the therapeutic interaction as it appears to the patient as well as to the therapist in retrospect. Seen in this perspective, the various statements, assertions, and attitudinal expressions by the patients refer less to specific behavior episodes "before," "during" or "following" therapy than they do to the manner in which the therapeutic experience has become integrated into the patient's total personality. The therapist's assessments are subject to the same considerations, although, because of his unique vantage point, they have perhaps a somewhat greater claim to "objectivity." In both cases, however, we are dealing with global impressions rather than with accounts of specific instances.

A further objective of this study was to explore the extent to which these general impressions can be broken down into more specific aspects of the therapeutic experience. Hence, our emphasis rested upon the isolation of "item clusters" which provided information concerning more *specific* aspects of the experience. The latter are exemplified by distinct, perceived qualities of the therapist's technique, such as his assumption of a relatively aloof and distant attitude, or his emergence as a person who may give reassurance, etc. We then proceeded to examine the statistical relationships among the various clusters for the patient and the therapist. Then we examined the consensus between patients' and therapists' assessments. Finally, some attention was paid to the relationships between attitudinal measures and such variables as the intensity of therapy and socioenvironmental status.

From a methodological standpoint, this investigation has given clues relative to the kinds of questions one might profitably ask former patients, the kinds of answers one may expect, and the constellation of items which might most successfully achieve this goal. As we have emphasized throughout, the present work has been largely exploratory, and we had to forge our measuring instruments as we went along. We could not draw upon scales whose characteristics had been thoroughly studied, simply because such scales do not exist. In the course of this investigation we have begun to isolate a number of relevant dimensions, but these must be developed further and strengthened in future research. Thus, we are not prepared to make assertions about the reliability (repeatability) of the various measures or their validity in terms of external criteria. On the other hand, we were able to develop a number of item clusters whose presumptive reliability and homogeneity is considered substantial. We advance this judgment on the basis of the usually high intercorrelations of items comprising a particular composite, which seem to lend each cluster considerable strength and stability. Further work, however, must be designed to produce purer measures of such composites as Therapist's warmth, Therapist's interest and respect, and others.

The therapist's clinical evaluations, too, are in need of clearer definitions and empirical anchoring points. One should not heavily rely on a presumptive consensus among therapists concerning variables like Anxiety, Disturbance, Ego strength, and so on. On the basis of the present results, however, future investigators can begin to work with specific dimensions that can be further developed and refined; they need not start with a vast mass of items (as we did) whose relevance to the research was largely intuited. In this sense the present work has made a methodological contribution.

Let us consider next the theoretical and practical significance of our substantive findings. The results and some of their implications have been commented upon as they were presented. At this point we wish to pull them together, realizing that some of our results may not be replicable. For this reason we shall largely confine ourselves to those findings in whose validity we have reasonable faith. Whether they are in fact replicable remains for future investigations to determine.

1. We were able to demonstrate the existence of a substantial retrospective consensus between patients and therapists concerning the essential aspects and outcome of the therapeutic experience. In this sense, even essentially subjective data are capable of a form of "consensual validation." Thus, if we assume that the therapist's evaluations contain an objective core, we can by inference, and on the basis of typical statistical relationships in the neighborhood of .60, attribute a measure of objectivity to the patient's assessments as well.

If the existence of a consensus can be accepted, the patients' detailed accounts (matched to some extent by the more meager therapists' judgments) provide a rich source of data concerning the manner in which therapy was experienced, the way in which the therapist and his technique were perceived, the nature of the problems which led the patient to enter therapy, and the character of the changes which the patients attributed to the therapeutic interaction. Personality changes resulting from psychotherapy have often been described by authors on the subject, but very rarely have such accountings been given by the patients. Considering the imperfections of the measuring instruments, the rather substantial agreement between patients and therapists in their view of each other and their assessment of the relationship and its outcome is particularly noteworthy and gratifying.

2. Evidence has been presented of certain differences in patient attitudes and experiences as a function of the form of therapy to which they were exposed. Whether the difference between psychoanalysis and psychotherapy is one of degree (as Alexander, 1956, has asserted) or one of kind (as a good many orthodox psychoanalysts believe) is a moot question and not answerable in terms of the present data, although it appears that the patients' experiences subtly shade into each other. To employ Tarachow's (1963) distinction, it is clear that patients in goal-limited or brief psychotherapy viewed their therapists considerably more as "real" objects than patients in more intensive psychotherapy (including analysis) where an "as if" experience predominated. This is not surprising and merely confirms what has long been held.

Not only was there a marked difference in the kind of relationship in which patients in different forms of psychotherapy were involved, but the data point to an important selection factor. Whether the patient was selected for intensive psychotherapy or more limited psychotherapy seemed to depend upon his age (young patients were preferred for intensive therapy), sex (male patients were preferred), socioeconomic status (more affluent patients were preferred), education (more highly educated patients were preferred). The selection of a patient for a particular form of therapy also seemed to depend upon such other factors as the patient's motivation to enter a prolonged therapeutic relationship, his degree of disturbance, the level of his anxiety and "inner pressure," his defensiveness, as well as other clinical considerations which were not fully reported on by the therapists. The further elucidation of the criteria for selecting patients for particular forms of psychotherapy, only generally hinted at in the data of this study, will mark a milestone in adapting different methods of psychotherapy to particular patients and their problems. It will also serve to place the enterprise of psychotherapy on a much more secure scientific footing besides conserving time, energy, money, and professional manpower.

Depending upon the form of psychotherapy, the patients' comments about their therapists varied to some extent. The most enthusiastic (and perhaps uncritical) reports stemmed from patients in the less intense forms of psychotherapy, whereas patients who had been treated by more intensive therapy tended to be somewhat more critical, analytical, and perhaps more realistic in their judgments. This may well have been a function of the degree to which the transference relationship had become a topic for analysis. Completely enthusiastic endorsements of the therapist seemed to come from patients who experienced the therapist as a "real object" and whose relationship never became the focus of interpretative scrutiny, whereas more tempered reports seemed to come from patients whose transference wishes apparently had not been gratified and who suffered real disappointment as a result. Whether they achieved more lasting therapeutic gains and underwent a more radical reconstruction of their personality cannot be determined from the data of our study. In general, however, more favorable evaluations of the therapist were a concomitant of positive therapeutic gains, as experienced by the patient, whereas an unsatisfactory therapeutic experience was found to be associated with negative attitudes toward the therapist, which apparently persisted well beyond the termination of therapy. This attitudinal variable seemed to be of crucial importance for the outcome, overshadowing differences in the therapist's closeness to or aloofness from the patient, depending upon the particular technique that was employed.

3. One of the most impressive segments of the results is represented by the patients' reports of the kinds of changes they attributed to the experience of psychotherapy. These commentaries by nonprofessional people, paralleling as they do the observations of therapists and theoreticians, are considered extremely valuable documentary evidence of an increased sense of ego mastery, the achievement of conscious executive control over impulses, symptoms, and neurotic trends, despite the often imperfect resolution of the neurotic character structure.

The patients' accounts also reveal that the removal of specific symptoms is only a minor aspect of the therapeutic gain they experienced. Even in the more limited and abbreviated courses of therapy there were gains which approximate the kinds of far-reaching changes experienced by patients undergoing more protracted therapy. This point is important to emphasize because it is frequently lost sight of by the proponents of those forms of therapy which have the reconditioning of symptoms as their objective. The significant gains which most of our respondents emphasize seem to concern preponderantly increments in ego mastery not only over specific symptoms but in a variety of areas of psychic functioning and adaptation to external reality. Thus, they forcefully call attention to the function of psychotherapy as an educational or re-educational process.

4. The patient's motivation for therapy has already been mentioned in the context of criteria for accepting patients for a particular form of psychotherapy and the characteristics of the therapeutic relationship resulting from this selection. A related problem concerns the motivational characteristics of patients who tend to derive the greatest benefit from psychotherapy. Our results suggest that the relationship is positive, and that patients describe themselves as having profited from therapy to the extent that they also report themselves as eager and willing actively to work on their problems, provided they are not too disturbed at the beginning of therapy.

5. With respect to the therapists' evaluations, we presented evidence to support several of our previous results, obtained both in experimental situations (psychotherapists responding to a patient presented by means of a sound film) and in initial diagnostic interviews: therapists derive greater enjoyment from treating patients with whom they established (or expect to establish) a productive working relationship; they prefer to work with patients whom they consider suitable for intensive psychotherapy, that is, those who have greater personality assets and potential resources, and in whose therapy they can invest themselves emotionally. This is

not to be considered synonymous with the assertion that therapists prefer to work with patients who are not seriously disturbed, but understandably they tend to select those who in their judgment possess appreciable potential for reconstructive therapy.

6. The emergence of a large general warmth factor in all forms of psychotherapy deserves particular emphasis. A solid working relationship in which the participants develop a sense of mutual trust is unquestionably a *sine qua non* for all forms of psychotherapy. In its absence there can be no successful psychotherapy. The presence of such a factor which suffuses and permeates all assessments, ratings, and evaluations, by patients as well as by therapists, was abundantly demonstrated. There was some additional evidence to suggest, however, that overshadowing this attitudinal-emotional factor is the patient's conviction that he has the therapist's *respect.* This faith in the integrity of the therapist as a person may be called the capstone of a successful therapeutic relationship subsuming all other characteristics. It should be reiterated that when a relationship can thus be described, both patient and therapist are firmly convinced of its existence. It is truly a shared feeling which admits of no doubt by the participants. Technical skill on the part of the therapist may go a long way to capitalize on such a relationship, and the data do not specifically inform us how such a relationship either comes into being, is deepened, or turned to maximum therapeutic advantage. However, there is little doubt a relationship having the qualities which we have described represents the most basic ingredient of beneficial therapeutic influence irrespective of the formal aspects of the setting.

SUMMARY

Objectives

This investigation was designed to provide preliminary data on the psychotherapy experience as perceived and evaluated by former patients in retrospect. The purpose was to gain a better understanding of factors in the patient-therapist relationship that might account for therapeutic change. Specifically, it was addressed to the following questions:

1. To what extent are particular treatment results associated with the manner in which the patient and therapist related to each other?

2. What part does the therapist's role behavior (personal style, techniques, etc.) play in treatment outcomes? Is it possible to identify attitude clusters that might differentially describe the therapeutic atmosphere?

3. To what extent do patients and therapists agree on (*a*) the quality of the working relationship; (*b*) their attitudes toward each other; and (*c*) treatment outcomes?

Subsidiary problems on which this investigation was intended to shed light included: the statistical association between length of treatment and quality of outcome, possible differences in therapeutic climate between "psychotherapy" and "psychoanalysis," the influence of socio-environmental variables on treatment results, the nature and quality of therapeutic change as reported by patients, residual problems, the patient's ability to cope with current problems, and the like.

Method

Two major instruments were employed: a patient and therapist questionnaire. Questionnaire P consisted of 89 items calling for descriptions of the patient's emotional difficulties before therapy, changes experienced, remaining problems and/or new ones, attitudes toward the therapist, the therapist's attitude as the patient experienced it, the kinds of therapist activity to which the patient attributed change, satisfaction with the results of treatment, etc. Questionnaire T was a brief rating instrument which to some extent paralleled the patients' assessments. It comprised 20 items dealing with global clinical estimates such as defensiveness, anxiety, capacity for insight, liking for patient, motivation for therapy, overall success of therapy, quality of the working relationship, etc.

Procedure

Eleven experienced therapists selected from their files a small sample of former patients, following certain broad criteria. Questionnaire P was mailed by each therapist to the patients in his sample. The therapist himself completed Questionnaire T for each of these patients.

Patient Sample

The original sample to whom questionnaires were sent comprised 78 patients, 39 men and 39 women. Of these, 44 were completed and returned by an equal number of men and women. The sample was considered comparable to the clientele of the average psychiatrist or psychoanalyst in private practice.

Results

A number of item clusters were isolated from Questionnaires P and T; in addition, certain individual items were retained. Coefficients of correlation were computed for the variables within each questionnaire and between questionnaires. Salient results may be summarized as follows:

1. We demonstrated the existence of a substantial retrospective consensus between patients and therapists concerning the essential aspects and outcome of the therapeutic experience. In this sense, even essentially subjective data are capable of a form of "consensual validation." Thus, if we assume that the therapist's evaluations are reasonably valid, we can by inference, and on the basis of typical statistical relationships in the neighborhood of .60, attribute a measure of objectivity to the patient's assessments as well. Personality changes resulting from psychotherapy have often been described by clinicians, but very rarely have such accountings been given by patients. Considering the imperfections of the measuring instruments, the rather substantial agreement between patients and therapists in their view of each other and their assessment of the relationship and its outcome is particularly noteworthy.

2. There were certain differences in patient attitudes and experiences which seemed to be attributable to the form of therapy the patients received. Patients in goal-limited therapy viewed their therapists considerably more as "real" objects, whereas in more intensive therapy an "as if" experience predominated. Furthermore, patients who were selected for intensive therapy tended to be younger, male, more affluent, and more highly educated. The patient's motivation to enter a prolonged therapeutic relationship and other clinical considerations of course influenced the selections as well. The further elucidation of these factors is considered an important task for the future.

3. Among the most impressive data were the patients' reports of the kinds of changes they attributed to the experience of psychotherapy. These commentaries document an increased sense of ego mastery, the achievement of conscious control over impulses, symptoms and neurotic trends, often despite the persistence of some neurotic problems. The removal of specific symptoms, in the patient's view, appeared to be a relatively minor gain from the therapeutic experience which he tended to describe as an educational or reeducational process.

4. Therapists derived greater enjoyment from treating patients with whom they established a productive working relationship, and they preferred to work with patients meeting the criteria for intensive psychotherapy. This finding is not tantamount to an assertion that therapists prefer to work with patients who are not seriously disturbed, but, understandably, they tend to select those who in their judgment possess appreciable potential for reconstructive therapy. On the other hand, patients described themselves as having profited from therapy to the extent that they seemed to have been eager and willing to work actively on their problems.

5. The emergence of a large general "warmth" factor deserves emphasis. There was additional evidence to suggest that overshadowing this attitudinal-emotional factor is the patient's conviction that he has the therapist's respect. This faith in the integrity of the therapist as a person may

be called the capstone of a successful therapeutic relationship subsuming other characteristics. Technical skill on the part of the therapist may go a long way to capitalize on such a relationship, although the present data do not specifically inform us how such a relationship comes into being, is deepened, and turned to maximum therapeutic advantage. However, there is little doubt that a relationship having these qualities represents the most basic ingredient of beneficial therapeutic influence irrespective of the formal aspects of the setting.

REFERENCES

ALEXANDER, F. *Psychoanalysis and psychotherapy.* New York: Norton, 1956.

BOARD, F. A. Patients' and physicians' judgments of outcome of psychotherapy in an outpatient clinic. *AMA Arch. gen. Psychiat.*, 1959, **1**, 185–196.

EYSENCK, H. J. The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961.

FEIFEL, H., & EELLS, JANET. Patients and therapists assess the same psychotherapy. *J. consult. Psychol.*, 1963, **27**, 310–318.

FRANK, J. D. *Persuasion and healing.* Baltimore: Johns Hopkins Press, 1961.

FRANK, J. D., GLIEDMAN, L. H., IMBER, S. D., STONE, A. R., & NASH, E. H. Patients' expectancies and relearning as factors determining improvement in psychotherapy. *Amer. J. Psychiat.*, 1959, **115**, 961–968.

GURIN, G. VEROFF, J. V., & FELD, SHEILA. *Americans view their mental health.* New York: Basic Books, 1960.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness.* New York: Wiley, 1958.

LORR, M. Relation of treatment frequency and duration to psychotherapeutic outcome. In H. H. Strupp, & L. Luborsky (Eds.), *Research in psychotherapy,* Vol. 2. Washington, D. C.: American Psychological Association, 1962. Pp. 134–141.

LORR, M., MCNAIR, D. M., MICHAUX, W. W., & RASKIN, A. Frequency of treatment and change in psychotherapy. *J. abnorm. soc. Psychol.*, 1962, **64**, 281–292.

LUBORSKY, L. The patient's personality and psychotherapeutic change. In H. H. Strupp, & L. Luborsky (Eds.), *Research in psychotherapy.* Vol. 2. Washington, D. C.: American Psychological Association, 1962. Pp. 115–133.

MALAN, D. H. *A study of brief psychotherapy.* Springfield, Ill.: Charles C Thomas, 1963.

MENNINGER, K. *Theory of psychoanalytic technique.* New York: Basic Books, 1958.

PFEFFER, A. Z. A procedure for the evaluation of the results of psychoanalysis. *J. Amer. Psychoanal. Ass.*, 1959, **7**, 418–445.

PFEFFER, A. Z. Follow-up study of a successful analysis. *J. Amer. Psychoanal. Ass.*, 1961, **9**, 698–718.

PFEFFER, A. Z. The meaning of the analyst after analysis. *J. Amer. Psychoanal. Ass.*, 1963, **11**, 229–244.

ROGERS, C. R., & DYMOND, ROSALIND, F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

SCHJELDERUP, H. Lasting effects of psychoanalytic treatment. *Psychiatry.* 1955, **18**, 109–133.

SNYDER, W. U. *The psychotherapy relationship.* New York: Macmillan, 1961.

STONE, A. R., FRANK, J. D., NASH, E. H., JR., & IMBER, S. D. An intensive five-year follow-up study of treated psychiatric outpatients. *J. nerv. ment. Dis.*, 1961, **133**, 410–422.

STRUPP, H. H. *Psychotherapists in action.* New York: Grune & Stratton, 1960.

STRUPP, H. H., WALLACH, M. S., JENKINS, JOAN W., & WOGAN, M. Psychotherapists' assessments of former patients. *J. nerv. ment. Dis.*, 1963, **137**, 222–230.

STRUPP, H. H., & WILLIAMS, JOAN V. Some determinants of clinical evaluations of different psychiatrists. *AMA Arch. gen. Psychiat.*, 1960, **2**, 434–440.

TARACHOW, S. *An introduction to psychotherapy.* New York: International Universities Press, 1963.

WALLACH, M. S., & STRUPP, H. H. Psychotherapists' clinical judgments and attitudes towards patients. *J. consult. Psychol.*, 1960, **24**, 316–323.

WALLERSTEIN, R. S., ROBBINS, L. L., SARGENT, HELEN D., & LUBORSKY, L. The psychotherapy research project of the Menninger Foundation: Rationale, method, and sample use. *Bull. Menninger Clin.*, 1956, **20**, 221–278.

(Received December 27, 1963)

APPENDIX

Case Histories

To give a clearer idea of the kinds of data emerging from individual records, as opposed to group trends and summary results to which major attention has been given so far, we have selected a few questionnaires for more detailed consideration as case histories. While the emphasis rests on the patient's account of his therapeutic experience and his previous and current life situation, we also present collateral information for the treating therapist. These case histories are fairly typical examples of the completed questionnaires that were returned to us; we believe the reader will agree that they are a tribute to the care, diligence, seriousness, and honesty of our respondents.

Case No. 0305

This patient, a young married woman aged 33, had been seen in psychotherapy for a period of approximately 3½ years, averaging 2 or 3 weekly interviews. At the time of answering the questionnaire, she was a housewife and portrait painter. She has two children. Family income was given as above $12,500 per year. Her education included 3 years of college. Her therapy had terminated about 2 years ago.

Her complaints that led her to enter therapy were described as follows:

> The term "worry yourself sick" applied to me. The weight of my many family illnesses and problems that were in areas where I was most vulnerable had led me to such repression of my worries that I was succumbing to a series of psychosomatic illnesses. I was in the hospital with Reynaud's syndrome when I decided things had gone far enough and entered therapy.
>
> I should explain that I had so internalized my problems that no one thought I needed treatment except me. Only at the very last did even my husband recognize that my problems were overwhelming me.

She saw herself as rather anxious and experienced intense internal "pressure" to enter therapy.

The therapeutic experience, for this patient, was exceedingly valuable, and she expresses herself as being "extremely satisfied" with the results. When asked for the reasons for her responses, she stated:

> Heavens above! I could write a book! My whole life has changed and I now live comfortably with myself and my family. Not only this, I have had a baby—now eight months old—and this has brought us all a joy beyond words.

The therapeutic changes were described in the following words:

> I am now at peace with myself. I no longer feel such a failure inside that I have to try to be 100% pluperfect on the outside and be all things to all men—so I can relax. I have discovered I'm more capable than I thought myself, for instance I have learned to paint and in fact paid for my treatment with money I made myself. I might add that I'm no longer quite so popular since I can now say "no" to my friends, but this doesn't bother me. I can pretty much take each day as it comes now and even derive much satisfaction from little things. Remaining problems include "factors that have nothing to do with me."

For example, she feels her relationship with her mother is still "painful" at times, about which she is sorry, but

> I also realize that it is something to be lived with as best I can.

This attitude, which has previously been described as a signal achievement on the part of many patients, does not appear to imply fatalism; rather it seems to signify a more realistic and discriminating attitude toward life and its problems. As one saying has it:

> God give me the power to change those things that can be changed, to accept those that cannot be changed, and the ability to tell the difference.

The patient felt that she was getting along very well now, and that her symptoms had completely disappeared. Interestingly, it was her impression that a fair amount of change has been apparent to her husband, but none at all to other persons.

This patient's responses are particularly revealing because she took the trouble to write amplifying comments on a number of the items dealing with her attitude toward the therapist and her therapeutic experience.

She thought he was interested in helping her, but made it clear that the amount of improvement she made was up to her. He was always polite, rather cool and impersonal, and very businesslike. Above all, she felt he would help her help herself.

She admitted to often feeling like "just another patient," but with this qualification:

> It was only after I could accept the fact that I *was* "just another patient" instead of someone who

had to be super special that I was on my way to being well.

A further comment on the therapist's attitude is illuminating:

> The therapist's attitude was rather stiff and formal—however, I took this to be his ordinary demeanor after a while, so I thought it perfectly natural. Some time after I completed therapy I met this doctor and his family socially and much to my surprise his personality was very unlike his professional personality. I have since decided that he varies his professional stance to best suit the personality of each patient.

Although she mildly disagreed with the statement that "Nothing the therapist said or did ever decreased my self-respect," she elaborated:

> I consider this one of the turning points of my treatment. Through an inadvertent slip one day, the therapist did say something that absolutely crushed me for a day or so. This happened as I was leaving and in the intervening two days between this visit and the next I grew up at least ten years. I learned that, though hurt, I still lived, that someone for whom I had great regard could make a mistake too and that I could get angry without the world coming to an end or lightning striking me dead! In my next visit, the therapist explained it was a slip and apologized (incidentally, the only time he ever made a reference to himself) and I further realized I could forgive someone for hurting me.

Concerning being considered a "worthwhile person":

> Like any patient I was concerned with his feelings toward me, but as time went on I was less interested in his liking me than I was in liking myself. I came to realize that it doesn't really matter how "worthwhile" you are in the conventional sense to a psychiatrist—if you are sincere and serious about your treatment he will do anything he can to help you even if you are the biggest louse there is in the community.

She attributed her therapeutic success, among other things, to the fact that she sought help by taking the initiative and by choosing her own therapist. Also, she recognized that her need for independence and her perfectionism (which in part were closely bound up with her neurotic illness) at the same time provided the motivation for seeking professional help and the persistence in working on her problem over an extended period of time. Finally, she cited as an ingredient in her cure the fact that she did not have a basic belief in a benign Deity. She believes that this is why many people do not seem to be able to go beyond a supportive type of psychotherapy because

> If one believes there is a God looking after one, it seems to leave a vacuum in the person when it comes to taking the basic responsibility for one's character and actions.

In conclusion, she viewed the fact that she was able to take this responsibility and turn to a professional person for help as

> a *real* miracle and certainly a lesson to me *never* to underestimate the human spirit.

The therapist, on his part, essentially agreed with the patient's favorable evaluation of her treatment experience, and termed the changes and the overall success "fairly great." His comment, after reading the patient's voluminous notes, was:

> Thank goodness we have a few results like this to keep morale up and give the feeling it's really worth doing. This patient is really a very remarkable person.

While the account is deficient in giving insight into the patient's psychodynamics and the process of psychotherapy, it strongly suggests that the results were at least partially due to impressive personality assets and resources which propelled the patient in the direction of a cure.

Case No. 0703

This patient, a 32-year-old single mathematician, who had earned the PhD degree and at the time of data collection was teaching at a university, had been seen for 200 hours in psychotherapy over a period of almost 2 years. Interviews averaged three sessions per week, and the patient's fee was $15 an hour. He had left psychotherapy 2 years prior to answering the questionnaire. Termination was occasioned by external factors (the therapist's going away on a sabbatical leave). The patient was moderately satisfied with the results of the treatment, credited himself with having obtained a certain amount of insight into some of his problems, but expressed the opinion that they were not worked out, which he attributed to the limited amount of time that was available. He stated that he had a need for further therapy, but he explained the need as not "urgent" which, coupled with the expense, kept him from seeking further treatment.

His presenting problem centered around a work inhibition although he was aware of other major difficulties including family relationships and sexual problems for which he hoped to obtain help. At the beginning, however, he carefully refrained from consciously acknowledging them as problems on the grounds that they would not be considered sufficiently important (presumably by the therapist) to warrant therapy. The therapist described this patient as rather schizoid, withdrawn, and aloof.

The patient considered his symptoms "somewhat improved" at the end of therapy and noted appreciable change after about 4 months of treatment (or about 50 hours). It appears that he developed a rather dependent relationship, which he described in these terms:

> My attitude toward him was somewhat similar (though not quite so strong) as that of a drowning person towards a log. Because of this, I purposely held back certain problems at the beginning of therapy and it took me quite a few months to persuade myself that the therapist was not on the verge—or would at least like to—kick me out of the therapy situation.

At termination, he acknowledged feeling a

> little angry at being deserted but at the same time I had the full realization of how irrational the feeling was—I was a little embarrassed by it.

His present attitude toward the therapist was described as "friendly."

Unlike most patients, he did not consider therapy an intensely emotional experience, although he regarded it as often rather painful. He experienced relatively little feeling in the course of his therapy and did not recall ever feeling intensely angry at the therapist at any time. He had a great deal of trust in the therapist's integrity, considered his attitude natural and unstudied as well as warm. He did not feel the therapist used technical language or that he engaged in small talk. He felt fully accepted and respected by the therapist. His self-respect was never decreased during therapy. Although he was convinced of the therapist's genuine interest in helping him, he had some question about the consistency of the therapist's attentiveness ("He was frequently hopelessly bored"). (Note a similar observation by Patient No. 0701 who was treated by the same psychiatrist.) It is possible that this therapist employed an attitude of studied boredom as a therapeutic technique. On the whole, the therapist was experienced as quite "active," and at times he seemed not averse to giving direct reassurance; still he gave no advice to the patient on how to conduct his life. There was the suggestion that in some respects this therapist emphasized intellectual understanding over the emotional aspects of the therapeutic experience. A comment by the patient may have some bearing on this:

> I feel that a valid and important distinction can be made between intellectual insight into a problem and an emotional working through of it. (The intellectual insight may well include the realization that the problem has not been emotionally worked through.)

He stated that he had gained *intellectual* insight into certain rather intense hostile feelings towards persons "having a safely peripheral role in my life." In addition, he noted a growing sense of being an integrated person ("an organic whole changing gradually through time rather than a series of disjointed snapshots having no integral relation to each other"). Other gains included: a reduction of feelings of guilt about not being properly affectionate towards members of his immediate family, an awareness that he was lonely, a realization that he could play an active part in changing certain situations if he so desired, and a firmer sense of his masculine identity.

His basic problems appeared unchanged, and he felt no change had become apparent to other persons in his life. Still, he saw himself as getting along "fairly well," and felt "somewhat more able" to cope with his current problems.

Although he found it difficult to answer the question of identification with the therapist's attitudes, values, or opinions, he recognized —"sometimes to my surprise"—that he still finds himself using occasionally a certain phrase or sentence of the therapist.

It is apparent that the patient himself approached therapy with a rather detached intellectual attitude, and the therapist commented in his evaluation about the patient's remoteness. He rated the patient's gains as moderate noting that "he warmed up a good deal." Although limited in scope, the therapeutic results in this instance appear to be commensurate with the particular character structure of this patient.

Case No. 0104

This case history is an example of a rather seriously disturbed girl who had been seen in psychoanalysis for a considerable period of time (5 years) by one of the most experienced analysts included in the sample. This patient had been hospitalized for an acute schizophrenic episode prior to entering therapy.

The patient was 18 years of age when she entered therapy in 1954. She is the daughter of rather wealthy parents who initially paid for her treatment. While in therapy she completed her college education and got married. She has one child. Her current family income is rather modest.

Her answers indicate that she had benefited from therapy to some extent although she considers herself in considerable need of further treament. Fairly satisfied with the results, she feels capable now of viewing her situation in a more mature manner. However, a great number of symptoms still persist, and she does not seem to be able to cope adequately with them.

Although she felt the therapist was genuinely interested in helping her, she often felt she was "just another patient." She considered the therapist to be stiff and formal, his attitude somewhat cold and distant, and she was never sure whether he considered her a worthwhile person. She had reservations about her trust in him, felt criticized at times, and recalled that the therapist tended to talk about psychoanalytic theory during her sessions. In her words:

> At the beginning I felt skeptical partially due to the procedure (saying "whatever comes to your mind") and the relative coldness (and to my mind thus indifference) of the analyst; also some feelings of hostility and resentment despite a great desire for help.
>
> At the end of therapy: Still somewhat mistrustful but I came to accept that this was due to a basic distrust of all persons; a far more friendly attitude prevailed.
>
> Now: Few feelings of hostility remain and there is a general feeling of friendliness although I tend to blame the analyst for not "helping me more."

She felt "somewhat undecided" about specific aspects of the therapist's activities which accounted for therapeutic change, but

> I do feel that his delving into my childhood experiences helped alleviate a number of unresolved guilt feelings; on the whole, however, I am not certain what aspect of his activities resulted in change.

She believed she identified with some of his attitudes and values:

> His more liberal attitudes toward sex and religion were largely adopted by me; his attitude toward my family and his opinion that my life was my own and should be led as I saw fit were also taken over by me.

The problems which led this patient to enter psychotherapy were described as follows: Severe depression; inability to get along with friends; extreme irritability; feelings of hopelessness and despair; unexplainable feelings of fear; very poor social adjustment. She viewed the disorder as chronic, thought she was getting along very poorly, felt rather anxious, and under extremely great "internal pressure."

In the light of her therapeutic experience she formulated the problem in these terms:

> My symptoms became meaningful; I had a great deal of understanding about my feelings and the reasons for them. However, I have since discovered the understanding was almost completely on an intellectual plane. Very little change (i.e., emotional) actually took place.

The gains from therapy, as already noted, appeared slight to the patient, but she admitted having achieved a more realistic attitude toward her problems; and her symptoms, while still existing, do not loom as large as before therapy, and they cause her less distress.

The therapist, in this instance, took a more favorable view of the outcome. He considered the overall success "fairly great" and expressed a similar judgment concerning the degree of basic characterologic change. He felt he had a "fairly good" working relationship with the patient and had a "moderate" emotional investment in her treatment.

After having examined her questionnaire responses he commented as follows:

> This patient had a severe paranoid schizophrenic disorder and was in the hospital prior to the commencement of psychotherapy with me. She made a limited improvement with reduction of guilt feelings and enhanced ability for reality-testing. At this time, she may be more amenable to a more psychoanalytic psychotherapy than was previously possible.
>
> I am impressed that the patient indicates I used "technical terms." She had courses in psychology at ____________ College and used some herself sometimes, but even so I think that what the patient regards as "technical terms" and what I would maybe were very much apart. Words were needed very much by this patient to name her feelings. There was, too, a difficult problem of silence for some periods, especially at the beginning of psychotherapy.

The success in this instance was only moderate, but in view of the severity of her disturbance the therapist's evaluation may be the more realistic one.

At the end of the questionnaire the patient mentioned that she had recently applied for further therapy to the mental hygiene clinic in her community and that she was currently seeing a psychologist in psychotherapy.

> I feel it is too early to try to evaluate the results of my visits,

were her closing remarks.

CASE NO. 0201

The patient is a 34-year-old married woman who has one child, and who is a high-school graduate. The family income was given as between $10,000 and $12,500 per year. At the beginning of therapy she was 30 years old.

Therapy began during her hospitalization, the reasons for which she does not give:

> I have never written down my symptoms. I feel a great reluctance toward doing so now—so I won't.

She was getting along very poorly then, felt very much disturbed and anxious, and felt a great deal of internal pressure, Following her hospitalization, during which period she was seen daily, she was treated by the same therapist for 1 year on a once-a-week basis in a face-to-face position. She paid $5.00 per hour. Although she did not seek further therapy thereafter, she felt in need of therapy a number of times; however, the distance involved from her home to Chapel Hill was too great. She and her therapist decided to try treatment closer to home, but this didn't prove satisfactory. Therapy was terminated by mutual agreement, but the patient sees herself as in considerable need of therapy now.

She felt fairly satisfied with the results of her therapy. As she put it: "I made progress but was not cured."

> My problems to me were the same at the beginning and end of therapy. My hospitalization relieved me, for a short period, of my domestic responsibilities, therefore making them at least tolerable.

She feels she has changed somewhat as a result of psychotherapy; however, she adds:

> I am no longer getting treatment of any description. This was my choice. I have been ill for such a long time now that I cannot remember my life as it was before. The only change I can describe at all is to tolerate the illness without getting hysterical.

The same problems that brought her to therapy are said to persist. Still, she believes she is dealing with them fairly adequately. A fair amount of change, such as she admitted, was apparent to her husband as well as to close friends and more remote associates.

She saw her therapy experience as intensely emotional and painful, and she remembers a good deal about it. She regarded the therapist as rather passive, stiff and formal, cold, and unaccepting. She often felt like "just another patient," complained about the therapist's inattention, and never felt sure whether the therapist considered her a worthwhile person. Neither was she convinced that the therapist was genuinely interested in helping her. Her attitude toward the therapist was summed up in the acerbic statement:

> I loathed him, personally; respected his ability.

The passage of 4 years had not mitigated the intensity of her negative attitude. In her words: "I still loathe him!" Elsewhere she said:

> Some days I needed reassurance. I'm sure I needed more than that but this is what I was aware of. My therapist was cordial, but never gave me any encouragement. My therapist's theory was *Silence*.

She denies having taken over any of his attitudes, values, or opinions. "He had no opinions," she commented wryly.

One surmises from the patient's responses that the therapist attempted to create a therapeutic relationship in accordance with the standard psychoanalytic model, which the patient (for reasons of her own) experienced as vacuous and unhelpful. Apparently, she was never able to accommodate herself to the therapist's technique and left treatment in a rather acute state of negative transference which never subsided. Although she felt benefited from therapy to some extent, she is rather bitter about the experience. It seems, too, that the patient, while not unintelligent, was rather obtuse in a psychological sense. She viewed her illness by close analogy to a medical one ("I cannot remember my life as it was before") from which she never recovered. In this connection, she asserted that she felt in need of professional help less than 1 year before seeking it.

Here is an example of a patient for whom psychotherapy was basically not a worthwhile experience, and one cannot help wondering whether some modification in technique might not have produced a more favorable result.

The therapist, for his part, rated the overall success of therapy with this patient as "very little." He felt he had a "fairly poor" working relationship with her, and characterized his overall experiences with this patient as predominantly unpleasant. He commented as follows:

Unfortunately this lady could not be seen more than once weekly. Each hour was isolated from the others. As soon as her defenses were threatened she became violently hostile and threatening. She had a severe obsessive-compulsive neurosis which covered vast reservoirs of hostility.

I had an opportunity to see her for one follow-up visit in the spring of 1961. She is unchanged—still using her symptoms to control mother, husband, and daughter, yet able to get out enough to enjoy life. As before, I recommended moving nearer to a source of therapy—as before she rejected this as impossible.

CASE No. 0203

A 32-year-old single male, this patient had completed 2 years of graduate study leading to the Master of Science degree in industrial relations and was on the verge of receiving the degree. He came from a fairly well-to-do family and was currently enrolled in a management training program for retail department stores.

This patient had been seen in intensive psychotherapy or psychoanalysis for almost 5 years, for a total of 400–450 hours. For the first 6 months of therapy, he was seen in face-to-face contact, thereafter reclining on the couch. The typical fee was $15 per hour. Eight months after termination, he returned for seven additional sessions with the therapist. At the time of completing the questionnaire, he felt no actual need for further therapy, but thought he could still benefit from therapy now and then. He expressed himself as extremely satisfied with the results of his therapy and felt he had benefited a great deal.

Prior to his coming to the university he had been seen by several therapists in supportive therapy. The last one encouraged him to seek more intensive therapy, which he did. His symptoms included: anxiety about relationships with women; anxiety about doing well academically: anxiety about relationships with parents and friends; obsessive-compulsive symptoms; lack of emotional tone; bodily aches and physical symptoms and many common colds. He considered his difficulties as having been of long standing (5–10 years), experienced extremely great internal pressure when entering therapy, a fair amount of anxiety, and thought he was getting along fairly poorly.

Marked change became apparent to the patient about midway in therapy, after 2 years. At termination, he felt his symptoms had greatly improved and he expressed the belief that they would completely disappear once he started working and got married.

Among the changes which he attributed to his therapeutic experience, he distinguished four areas: (*a*) an increase in emotional tone, (*b*) insight, (*c*) understanding of others, and (*d*) external changes. He elaborated on these in the following terms:

> I am now able to have "pure" strong feelings. By pure I mean comfortable feelings. Thus, when I was sick and *thought* I was happy at times it felt more like a nervous exhilaration or manic state rather than the feelings I have now which are truly comfortable and joyous and relaxing feelings. Likewise, now when I get angry I have a pure, strong feeling of anger. Expression of this anger brings relief. Anger makes me aggressive and makes me look for a solution and catharsis. Formerly, anger made me nervous, frightened, submissive, and eventually very depressed.

He has become able to be his own therapist and to handle formerly frightening or inexplicable feelings or phantasies. Others previously seemed mysterious people, who seemed to know more about the patient than he knew about himself. Now, he feels he has an appreciation of the problems, phantasies, and motivations of others after having learned a great deal about himself. He feels he has many fewer colds, sleeps much better, and has more physical energy.

He still suffers from some lack of confidence in his own abilities, but he knows that he can succeed when he gets into the swing of things. Upon entering college, he felt he would "flunk out," but he made Phi Beta Kappa and went on to do graduate work.

He believes a great deal of change has been apparent in close associates, but to co-workers and associates the changes have been apparent only "on an unconscious level."

The patient experienced therapy as an intensely emotional experience (he commented that this particular question sounded like one aimed at Park Avenue millionnaires). He felt the therapist was genuinely interested in helping him; he never felt like "just another patient;" he experienced the therapist as understanding and warm; and had many positive and complimentary things to say about the therapist's empathy, unequivocal honesty, "shocklessness," and his unflagging attention. He commented on the therapist's ability to summarize trends together, without ever getting too far ahead of the patient. He was clearly aware of having identified with many of the therapist's positive attributes, his love for truth, and his desire to understand people and their motivations. He developed a great liking for the therapist, and feels "a life-long bond" with him; he admired his integrity, genuine desire to help people, his considerateness, and sense of humor; and he has "complete confidence in him and no feelings of hesitation about verbalizing or expressing any feelings or thoughts."

In retrospect, he formulated his problem as follows:

> Repressions (beginning about age 2½) which were heavily reinforced until about age 19 when I was first hospitalized. I had "learned" to deny most all my feelings. All my energies which "should have been used" in a healthy, realistic way were being bound up repressing feelings of hostility, jealousy, lust, attraction, etc., which I was unable to accept or verbalize. Repressed feelings "caused" lack of emotional tone. Repressed hostility "caused" depressions. Therapy enabled expression, verbalization and acceptance of childhood fears, emotions, phantasies.

The patient's unbounded (?) trust in the therapist is in striking contrast to the negative comments by Patient No. 0201 about the same therapist. In rating this patient, the therapist described him as extremely anxious at the beginning of therapy, expected some improvement, but noted fairly great symptomatic improvement upon termination. Yet he felt there had been only some change in the patient's basic personality structure. He felt that he had a fairly great "emotional investment" in the patient and described the working relationship as "fairly good." The overall success of the therapy was rated by the therapist as "moderate." Beginning as largely supportive therapy, it became intensive-analytical. He recalled both strikingly pleasant and strikingly unpleasant experiences with this patient. His laconic comment was:

> In 5 years of intensive therapy of a schizophrenic one has all sorts of pleasant and unpleasant experiences!

CASE NO. 0701

The patient in this case is a 28-year-old married male, without children, who was seen on a fairly intensive basis (three times a week) for a relatively short period of time (about 1 year). Like many patients represented in the sample, this respondent was highly educated—he had attended graduate school, without obtaining a degree—and was currently occupying a managerial position in a contracting firm. His income was given as between $10,000 and $12,000. In therapy, both the couch and the face-to-face position were used at one time or another. The therapist described the form of therapy as gravitating toward the intensive-analytical pole.

> The symptoms which had begun 3–4 years prior to therapy included a sort of claustrophobia, especially in crowded places—I noticed that the intensity of these symptoms depended on how interested I was in something else, e.g., very little trouble at an exciting ball game, a lot of trouble during a dull sermon. Therefore, I was sure something could be done about it with competent professional help. The worst problem was that although I knew pretty well what was wrong, all logical methods of fighting it made it worse. It was like fighting a ghost—I knew there were no ghosts yet one had me on the run. Also, there were occasional migraine headaches (5–6 a year).

This patient appeared to be under somewhat less strain than many other respondents when they entered therapy.

He described himself as being extremely satisfied with the results of his treatment, felt no need for further therapeutic work at all, and thought he had changed a fair amount. Commenting on the nature of the changes, he noted that only those closest to him would be able to notice, but he considered this as perhaps the greatest change of all:

> I have learned to accept myself as I am, and I find that I rather enjoy being me; the most drastic change, in other words, has been to learn that I can live an enjoyable life without any drastic changes. I do not mean that I now make no effort to improve myself; on the contrary, I think I am more intent on this than before. The difference is in my outlook and attitude more than anything else. As a result of the above, I find that I have more self-confidence, and am willing and able to tackle things that I would not have attempted before, though I may not like it any better.

He believes that he is getting along very well now and he faces no problems that he cannot cope with in one way or another.

This patient, at least in retrospect, had a fair amount of confidence in his therapist from the beginning.

> I was looking for help and he was "it," as far as I was concerned.

He felt this in itself was a very important ingredient in his progress.

At the time of termination, the therapist recommended some further treatment, but the patient thought that the therapist

> had done a better job than perhaps he realized, for I believe that I was far enough along to where I could work things out myself. I felt that my con-

fidence in him was justified and that he had helped me a great deal.

This patient's comments, too, provide some evidence that therapy does not "stop" with the last interview, rather that the process which was initiated during the meetings with the therapist continues often for a considerable period of time:

> I feel that I am still improving as I go along, by applying what I learned while under treatment.

Although this patient was seen in intensive psychotherapy, the therapist did not follow the psychoanalytic model rigorously. True, the patient admitted feeling intense anger toward the therapist at times, but nothing the therapist said or did ever decreased his self-respect. He felt fully accepted as a worthwhile person, perceived the therapist's attitude as warm, and considered the therapist's interest as genuine. The focus in therapy rested on the patient's relationships with people in his current life, and one surmises that the therapist was less concerned with fostering the development of a transference neurosis or in interpreting the transference. Still, his technique was essentially interpretive, as the patient reported it, and he fully left the initiative with the patient:

> His attitude showed that he was genuinely interested in helping me, but at the same time made it clear that most of it was up to me—he was neither overly friendly nor overly aloof—he was bored when the occasion would likely have bored most people, and somewhat enthusiastic at times, also.

In sum, it appears that the therapist emerged more clearly as a "real" person in therapy with this patient.

The therapist shared the patient's evaluation and considered him markedly improved. Here, then, is an example of analytically oriented psychotherapy which bore fruit, although one may question that it reached the deepest layers of the patient's personality and effected pervasive characterologic changes. The symptomatic improvement, however, is impressive.

Vol. 78, No. 12 Whole No. 589, 1964
Psychological Monographs: General and Applied

VISUAL PERCEPTION IN STRABISMUS AND AMBLYOPIA[1]

SAMUEL C. MCLAUGHLIN
Institute for Psychological Research, Tufts University

Strabismus is a visual disorder in which one eye is deviated from the line of sight. Amblyopia (more properly, amblyopia ex anopsia, which means "deterioration of vision through disuse") is a partial deficiency of visual acuity in 1 eye—usually a strabismic eye. Evidence is presented that the underlying disorder in the strabismus-amblyopia syndrome is a previously unrecognized anomaly of visual perception: objects as seen from 1 eye are perceived as being unreal. The term "unreal" in this context means "incapable of stimulating any sense modality except the visual." Operations appropriate to this definition are described. The perceptual disorder, once acquired, leads to a progressive series of changes in the appearance of the unreal image, the end result being "suppression" of that image. This is a perceptual phenomenon: the entire process takes place in the absence of any change in the eye itself. Strabismus therefore provides unique opportunities for studying the action of the visual interpretative mechanism in normal seeing. Evidence is presented that the perceptual disorder in question accounts for the major clinical features of strabismus and amblyopia. Implications for the treatment of these disorders, and for the understanding of normal binocular coordination, are discussed.

STRABISMUS is a visual disorder in which one eye is deviated from the line of sight. The deviated eye may turn inward (toward the nose) or outward (away from the nose). Amblyopia (more properly, *amblyopia ex anopsia*, which means deterioration of vision through lack of use) is a partial deficiency of visual acuity in one eye—usually in a strabismic eye. Unlike other disorders which reduce visual acuity, amblyopia ex anopsia is not associated with any physiological change in the eye itself.

It has long been recognized that strabismus and amblyopia are of interest for the study of visual perception. In a recent symposium on sensory deprivation Teuber (1961) pointed out that in strabismus

> we are likely to see curious things important for a theory of perception.... It seems to me [he said] that the perceptual deficit in the amblyopic eye is quite analogous to the impaired vision of Dr. Riesen's cats [p. 34].

Commenting on Teuber's remarks, D. O. Hebb (1961) added,

> it seems incredible that there has been no psychological research [in this field].... The ophthalmologists are all familiar with such problems, but psychologists seem not to have heard of them [p. 43].

During the past 6 years, the present writer has been investigating the visual behavior of strabismic and amblyopic patients. Some of the results of this research, having to do with the etiology and treatment of the strabismus-amblyopia syndrome, have been reported in the medical literature (McLaughlin, 1959a, 1959b, 1960, 1964). The purpose of the present paper is to describe some observations and findings which may be of interest to psychologists.

STRABISMUS-AMBLYOPIA SYNDROME

Nonparalytic Strabismus

Paralytic strabismus is a condition in which the patient cannot hold his deviated eye straight under any circumstances. The paralysis may have a neurological origin, or there may be damage to the oculorotary muscles themselves.

---

[1] This research was supported in part by Public Health Service Research Grants B-3598, B-3867, NB-04540-01, and NB-04540-02, from the National Institutes of Health, Division of Neurological Diseases and Blindness.

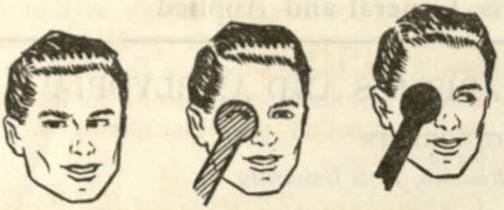


FIG. 1. Left: a strabismic individual looking at a distant object directly in front of him. (His left eye is deviated inward.)

Center: the same individual, still looking at the same object, after his right eye has been covered. (His left eye has straightened, thus demonstrating that he has the nonparalytic type of strabismus. The cover is opaque; the cross-hatching is intended to show the patient's right eye turning in behind the cover.)

Right: the patient looking at an object off to his left, demonstrating that his right eye has little or no restriction in range of movement.

Contrary to widespread belief, paralytic strabismus is the exception rather than the rule among strabismic patients. The great majority of strabismic patients have what is known as "nonparalytic" or "concomitant" strabismus. (The two terms are alternative names for the same condition.) A patient with this type of strabismus is able to hold his strabismic eye straight when his other eye is covered (see Figure 1). This does not mean that he then holds both eyes straight at the same time, for the eye behind the cover turns in, as shown in the middle drawing of Figure 1. That is, the two eyes undergo a movement which is more or less concomitant. Ordinarily, the strabismic eye—the eye that is usually deviated—shows little or no restriction in range of movement (right-hand drawing in Figure 1).

Absence of Visual Handicap

A person with normal binocular vision tends to see double whenever his eyes are not properly aligned. The reader may be able to experience this phenomenon by pushing one eye out of alignment or by deliberately overconverging his eyes (see Figure 2). If a person were to have one eye continuously out of alignment, he might be expected to have a serious visual handicap. On looking at a pencil, for example, he would see two pencils; and he might have some difficulty in deciding which one could be used to write with.

When an individual has only recently become strabismic, he does see two of whatever object he looks at, and he cannot easily decide which one to reach for. But an individual with a well-established strabismic condition no longer has this problem. Typically, such a person has no awareness of "double vision" and suffers no visual confusion. This absence of visual handicap is characteristic of nonparalytic strabismus.

Course of Onset of Nonparalytic Strabismus

In order to account for these remarkable characteristics of concomitant strabismus, it seems necessary to postulate a three-step sequence of events in the onset of the disorder.

Precipitating Factor. The initial cause of onset is a physiological defect in the visual mechanism which makes it difficult

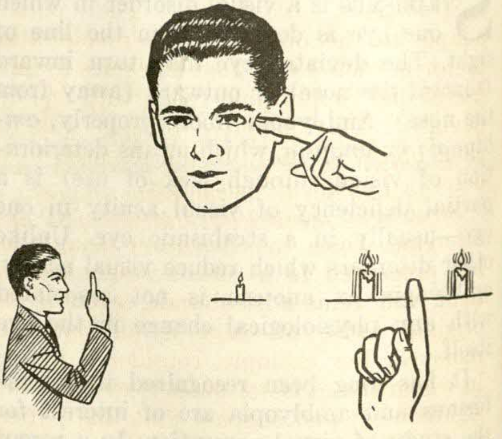


FIG. 2. Top: by applying light, intermittent finger pressure near the outer canthus of one eye, so as to displace the eyeball momentarily, a person with normal binocular vision can sometimes become momentarily aware of double vision. (The demonstration is more effective if the test object is a small high-contrast object, such as a candle in a darkened room. Holding a piece of colored transparent plastic over one eye also helps.)

Bottom: another way of experiencing double vision, namely, by looking at some near object (such as your finger). (The two images of some distant object, such as a candle, may then be seen in two different places.)

or impossible for the patient to use his eyes together in the normal fashion. For example, if an infant is unusually farsighted, he must exert an excessive amount of accommodation (and hence an excessive amount of convergence) in order to look at a near object. As a result, such an infant may develop a convergent strabismus.

Any difference between the two eyes, whether congenital or acquired, brings about a predisposition toward strabismus. For example, an individual may be farsighted in one eye and nearsighted in the other; or one eye may be slightly damaged, either at birth or later.

In many strabismic patients it is not possible to discover any defect in the visual system which might have caused the strabismus. To account for these cases, it has been postulated (Keiner, 1951) that myelinization of the oculomotor fibers is sometimes delayed beyond the usual period, producing a temporary oculomotor disability at the age when binocular coordination would normally develop.

Strabismic Adaptation. The patient adapts to the presence of a precipitating factor in such a way that he comes to have useful and comfortable vision with one eye deviated. This adaptive process was the chief object of study in the research to be described in the present paper.

Strabismic Adaptation Becomes Self-Reinforcing. Once the strabismic adaptation has proceeded to a certain stage, the patient will remain strabismic, even if the original cause of onset is alleviated or removed. For example, consider a patient in whom the cause of onset is extreme farsightedness; and suppose that this patient is provided with corrective lenses when he is 7 years of age. It is a common clinical observation that this is not enough; removal of the original cause of onset at this age will seldom remedy the strabismus.

For another example, consider a patient in whom the cause of onset was an anatomical deformity of the muscles which position the eyes; and suppose that this condition is corrected surgically when the patient is 7 years old. The same generalization applies: after surgery, the patient's eyes may be straight enough so that he no longer appears "cross-eyed" or "walleyed"; but the chances are that he would still be a suitable experimental subject for an investigation of visual perception in strabismus.

Contemporary Theories about the Strabismic Adaptation

There are three views concerning the strabismic adaptation which have gained widespread acceptance: suppression, abnormal retinal correspondence, and anomalous projection.

Suppression. In the contemporary medical literature, the strabismic adaptation is perhaps most frequently described by the statement that the patient comes to "suppress" visual information from his deviated eye. This is a valid and useful concept. To a considerable extent, the study of the strabismic adaptation is the study of monocular suppression.

Retinal Correspondence: Normal and Abnormal. Some strabismic patients—notably those who acquire strabismus in early infancy—seem to avoid double vision without suppressing information from the deviated eye. This is illustrated in Figure 3, which shows a patient with nonparalytic strabismus (left eye deviated inward) looking at a coffee cup. It is assumed that the patient has no awareness of double vision, and that he has 20/20 vision in each eye. The demonstration is carried out under ordinary room illumination, and the viewing distance is 3 meters. The patient's angle of strabismus, as measured by prism and cover (Kramer, 1953, p. 224), is assumed to be 20 prism diopters.

While the patient continues to look at the coffee cup, the experimenter places a small test letter 60 centimeters to the right of the coffee cup, as shown in the figure, so that the letter is imaged on the fovea of the patient's left eye. Even though the patient continues to look at the coffee cup, he can see the test letter clearly and can correctly identify it. He cannot do this with his left eye covered. In other words, the patient has two areas of clear vision in his binocular visual field. There are ob-

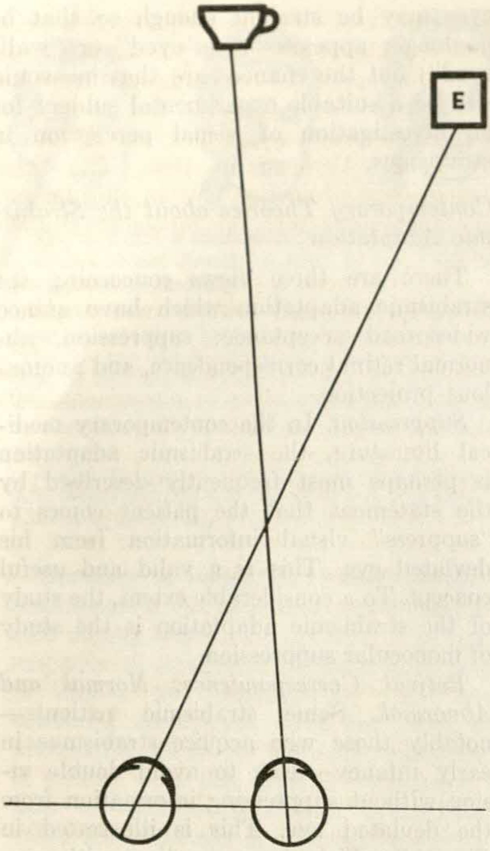


FIG. 3. Viewing situation which demonstrates two areas of clear vision. (The observer is assumed to have nonparalytic strabismus with his left eye deviated inward. He is instructed to look at the coffee cup, and does so by directing his right eye toward it. The test letter E is then positioned so that it is imaged on the fovea of his left eye. The patient does not see double, but absence of left-eye suppression is suggested by the fact that he sees the test letter clearly and can correctly report its orientation.)

vious difficulties in assuming that such a patient suppresses information from his deviated eye.

This phenomenon and others which appear to be incompatible with the notion of monocular suppression are frequently explained by assuming that the patient has "abnormal retinal correspondence." In order to understand this assumption, it is necessary first to consider normal retinal correspondence.

As already noted, a person with normal binocular vision tends to see double whenever his eyes are not properly aligned; and this double vision resolves into single vision when the eyes become realigned. There seems to be no way to explain this phenomenon except by assuming that, for each point in the retina of the right eye, there is a "corresponding" point in the retina of the left eye, such that two corresponding points have a common visual direction in external space. According to E. G. Boring (1942, p. 227), this concept of "retinal correspondence" dates back to Galen (ca. 175 A. D.). Isaac Newton, in 1730 (1952, p. 346), made it a fairly explicit neuroanatomical hypothesis. Retinal correspondence is "normal" when the fovea of the right eye and the fovea of the left eye are corresponding points.

A diagnosis of abnormal retinal correspondence is equivalent to an assumption that the fovea of each eye, instead of corresponding to the fovea of the other eye, corresponds to some nonfoveal point (or area) of the other eye. This assumption further states that, although intraretinal relationships are unchanged, the entire retinal mosaic of the right eye is displaced horizontally in terms of its correspondence with the retina of the left eye, the magnitude and direction of the displacement being appropriate to the patient's angle of strabismus. (For an outline of the history of this concept, see Duke-Elder, 1934, pp. 1029–1030.)

There are two different ways of interpreting the concept of abnormal retinal correspondence. The empiricist view (Helmholtz, 1925, pp. 409, 539; Duke-Elder, 1934, pp. 1029–1030) is that normal and abnormal correspondence are equal valued, that neither is innate, and that a given individual may develop one or the other, depending largely on whether or not he is required to adapt to a strabismic deviation. The nativistic view is that normal retinal correspondence is innate and unalterable (see, for example, Linkz, 1952, p. 548). According to the nativistic view, abnormal correspondence is a superficial and unstable condition which relies for its existence on the continued repression of normal correspondence.

There are several reasons for favoring the nativistic view. For one thing, the differences between normal and abnormal correspondence which are predicted by the nativistic theory do in fact exist. As F. H. Verhoeff showed in 1938 (p. 666), even the defining characteristic of retinal correspondence (common visual direction for corresponding points) is missing from the abnormal type of correspondence.

Another reason for favoring the nativistic view has to do with the awareness of double vision in strabismic patients. According to the empiricist theory, it should be impossible to arouse an awareness of double vision in a patient who has abnormal retinal correspondence except by placing a wedge prism before one eye. Without a prism, if the patient can be made simultaneously aware of the two monocular representations of a single object, he should see them superimposed—i.e., abnormally corresponding. It has recently been shown that, under appropriate conditions of viewing, patients previously diagnosed as having abnormal retinal correspondence can in fact be made aware of double vision without the aid of a prism (McLaughlin, 1959b, p. 84). This finding (which will be presented in more detail in a later section of this paper) indicates that, even in abnormal retinal correspondence, normal correspondence is intact but latent.

In view of these findings, it will be assumed here that the various phenomena which have been interpreted by other writers as abnormal retinal correspondence represent the action of adaptive perceptual mechanisms which intervene between (*a*) the basic neuroanatomical structure of (normal) retinal correspondence and (*b*) conscious visual perception.

Anomalous Projection. There is a way of interpreting the demonstration in Figure 3 (single vision but no suppression) without assuming any abnormality of retinal correspondence. To do this, it is only necessary to assume that the anomaly is one of visual localization rather than one of retinal correspondence. According to this view, the strabismic patient adapts to his condition by relocalizing the second image. The end result of this adaptive process is that any object seen from the deviated eye is localized as if it were being seen from the nondeviated eye. This is the theory of anomalous projection, which was advanced by F. H. Verhoeff in 1938. The term "anomalous projection" is an abridgment of "anomalous form of binocular projection."

The differences between these two theories—abnormal retinal correspondence and anomalous projection—are largely epistemological: both theories describe and predict the same class of phenomena, but the theory of anomalous projection does so with greater internal consistency and with more conservative assumptions. For this reason, the theory of anomalous projection provides the better starting point for further analysis of visual mechanisms in strabismus.

Amblyopia

There are two factors which contribute to the deficiency of visual acuity in amblyopia ex anopsia. These are (*a*) eccentric fixation and (*b*) suppression amblyopia. Contemporary theories about the strabismus-amblyopia syndrome deal with these two factors separately.

Eccentric Fixation. In terms of rotational position of an eye, normal fixation may be defined as positioning of the eye so that the object of regard is imaged on the fovea centralis. Eccentric fixation is a condition in which the object of regard is imaged on some nonfoveal area. Since the nonfoveal portions of the retina have lower visual acuity than the fovea, eccentric fixation brings about a reduction in visual acuity.

As shown in Figure 1, it is typical of concomitant strabismus that the deviated eye straightens when the other eye is covered. In some cases of strabismic amblyopia, however, the amblyopic eye remains partially or wholly deviated even when the other eye is covered, and even though the patient continues to "look at" the test object. This is the type of eccentric fixation which is frequently associated with amblyopia ex anopsia. The departure from normal fixation may be so great that the eye remains completely deviated or so small as to be detectable only by special ophthalmoscopic devices.

A patient with gross eccentric fixation may respond to the cover test (see Figure 1) in much the same way as a patient with paralytic strabismus. However, a patient with eccentric fixation will prove able to straighten his amblyopic eye under certain conditions of viewing, whereas the patient with paralytic strabismus will be unable to straighten the eye under any conditions of viewing.

It is generally agreed that eccentric fixation, as it occurs in strabismic amblyopia, is secondary to the strabismic deviation. The eye comes to fixate eccentrically because it has habitually been used that way. Eccentric fixation appears to be an instance of a binocular "habit" carrying over to monocular viewing.

Suppression Amblyopia. Suppression amblyopia may be defined as the deficit in visual acuity in amblyopia ex anopsia which remains after eccentric fixation has been eliminated, or which cannot be accounted for by eccentric fixation. This is the phenomenon which Teuber likened to the impaired vision of Riesen's cats. A patient with suppression amblyopia may have poor visual acuity despite the presence of a clear retinal image and accurate foveal fixation.

It is frequently stated in the medical literature that this type of amblyopia derives from monocular suppression in much the same way that eccentric fixation derives from the strabismic deviation (see, for example, Henderson, 1962, p. 101). According to this view, suppression amblyopia, like eccentric fixation, starts as a binocular anomaly and eventually carries over to monocular viewing.

ADAPTATION TO PARALYTIC STRABISMUS

The theories of abnormal retinal correspondence and anomalous projection appear to be based exclusively on observations and experiments involving patients in whom the strabismic adaptation had become well-established. In 1956, when the present investigation started, it was decided to try a different approach. Instead of observing only the end result of the adaptive process, it was decided to try to observe the successive stages of that process. This technique—observation of the course of onset of the sensory-motor adaptation—was suggested by the work of G. M. Stratton (1896).

Experimental Subject

The orthoptic staff of the Kresge Eye Institute in Detroit, who contributed in many ways to this research, provided an experimental patient who was an ideal subject for the study of the strabismic adaptation. This was a man about 30 years of age. In the course of his duties as a municipal employee, he had been struck on the side of the head with a sharp instrument in such a way that the muscles which controlled the position of one eye had been damaged. As a result of this injury he had a paralytic strabismus with the left eye deviated inward. There were no neurological complications, and his deviated eye was sound and functional.

Real-Unreal Distinction

Immediately following his accident, and for at least 2 years thereafter, the patient saw double. Within a few days after the accident, however, he had begun to make a distinction between "the real one" (seen from his nondeviated eye) and "the unreal one" (seen from his deviated eye). As his adaptation progressed, objects as seen from his deviated eye became more and more unreal. One year after the accident this real-unreal distinction was so intense that, despite his continued awareness of double vision, the patient's visual handicap was greatly reduced. He could, for example, drive an automobile through city traffic with both eyes open and uncovered. Two years after his accident, the real-unreal distinction had become even more intense: by this time, the second image was "so unreal that sometimes I don't even see it."

These observations show a remarkably close agreement with Stratton's report of his own visual experience when adapting to a 180° disorientation of his visual environment. Stratton (1896, p. 613) reported that the first stage of his adaptation appeared to be a perception of the reoriented images as unreal. The floor, for example, appeared overhead; but it appeared unreal there, so that it could not be mistaken for a floor on which one might walk. Evidently, the loss of perceptual reality is a key factor in adaptations of this sort.

Change in Appearance of the Second Image

A major thesis of the present paper is that the strabismus-amblyopia syndrome provides a powerful investigative tool for the study of normal visual perception. Nothing illustrates this point more cogently than the progressive changes in the appearance of the second image during the course of the strabismic adaptation. These changes are illustrated in Figure 4, which shows the two monocular representations of a doorknob as they appeared to the patient at each of several stages of the adaptive process.

For a month or so following the accident, the two images of the doorknob were indistinguishable. It seemed to the patient, in fact, that this was the chief source of his difficulties: the two looked so much alike that he sometimes did not know which one to reach for in order to open the door.

By the end of the second month, the patient could make correct motor responses if he were given a second or two in which to make up his mind about "which one to reach for." At about this time, he began to describe the second image as appearing "two dimensional," "not solid," "unsteady," although these characteristics of the image were not always evident to him at a glance. The second stage in Figure 4 is intended to illustrate this state of affairs.

Eighteen months after his accident, the patient was still aware of double vision; but he had little or no visual confusion. At this time, the second image appeared to him to be transparent, a "ghost image," "floating in space," "not really there" (third stage in Figure 4).

There was no abrupt transition from one stage to the next, and there were frequent regressions to an earlier stage. Even the disappearance of the second image (fourth stage) was not a clear-cut event. Two years after his accident, when the patient was reporting that he seldom saw double, he was still aware of the second image whenever he was asked about it.

Throughout this entire adaptive process, there was no change in the patient's deviated eye which might have accounted for these changes in the appearance of things seen from that eye. In fact, all of the anomalous characteristics of the second image vanished, and an object seen from his deviated eye appeared opaque, solid,

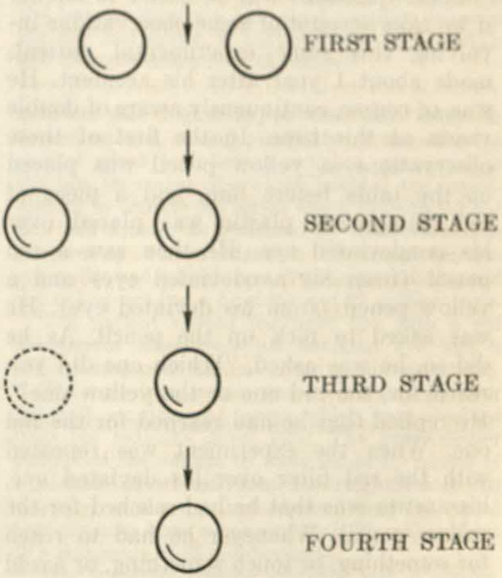


Fig. 4. Each of the four drawings shows the two monocular representations of a doorknob, as they appear to a patient during the course of the strabismic adaptation. (In each drawing, the small vertical arrow—not seen by the patient—represents the actual position of the doorknob. The four stages cover a period of about 2 years following the onset of the strabismus.)

(In the uppermost drawing [first stage] the patient has only recently become strabismic. He makes no distinction between the two images, and may not know which one to reach for in order to open the door. In the second stage, one image is correctly localized. Given time to make a conscious judgment, the patient can choose the correct image. In the third stage, one image appears to be real, the other appears unreal: the patient now has little or no visual handicap. The fourth stage is suppression of the unreal image.)

and tangible as soon as his other eye was covered.

Loss of Nonvisual Qualities

William James said in 1890,

> whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes ... out of our head [1950, p. 103].

Clearly, the change in the strabismic image represented a loss in the part that came out of the patient's head, but exactly what component, and why was this particular component missing?

These questions will be easier to answer if we take account of some observations involving this same experimental patient, made about 1 year after his accident. He was, of course, continuously aware of double vision at this time. In the first of these observations, a yellow pencil was placed on the table before him, and a piece of red transparent plastic was placed over his nondeviated eye. He then saw a red pencil (from his nondeviated eye) and a yellow pencil (from his deviated eye). He was asked to pick up the pencil. As he did so, he was asked, "Which one did you reach for, the red one or the yellow one?" He replied that he had reached for the red one. When the experiment was repeated with the red filter over his deviated eye, his answer was that he had reached for the yellow pencil. Whenever he had to reach for something, or touch something, or avoid bumping into something, his motor activity was invariably oriented toward the object as seen from his nondeviated eye. When he was asked why this was so, he replied that objects as seen from his deviated eye did not look like things that could be touched or manipulated or bumped into.

In the next observation, the test object (seen double) was a bell, and again the two monocular images were differentiated by placing a red filter before one eye. The bell rang, and the patient was asked which bell the sound came from. His reply indicated that the bell that rang was the one seen from his nondeviated eye. The bell seen from his deviated eye appeared to him to remain silent. The question of accuracy of auditory localization in this situation is completely irrelevant. Only one of the two bells looked as if it made a noise, and the other one looked like something that did not and could not make any noise.

Next, it was a rose that was seen double, and the patient was asked which one had an odor. He replied that it was the rose as seen from his nondeviated eye that had an odor. The other one, he said, was odorless. (Again, the question of accuracy of olfactory localization is not germane.)

When it was the flame of a candle that was seen double, only one image (the one seen from the patient's nondeviated eye) appeared to give off heat. The second image of the candle (seen from his deviated eye) did not even appear warm, any more than a picture of a candle appears warm.

These observations show what is missing from strabismic images. As the adaptation progresses, they gradually come to look like things that cannot be touched, cannot be heard, cannot be smelled, cannot be experienced in any way except through the sense of sight. In short, they come to lose all nonvisual qualities.

Relation between Change in Appearance and Loss of Nonvisual Qualities

The change in appearance of the second image is nothing more than the externalization of the loss of nonvisual qualities. So long as the two images were indistinguishable, the patient had to make a conscious judgment, prior to every step and prior to every gesture, as to "which one to reach for" or "which one to avoid." But the change in appearance of the second image gradually took these decisions out of the realm of conscious deliberation and made them matters of immediate visual experience. To say that the second image appears two dimensional, ghostlike, or unreal is to say that the quality of intangibility appears to reside in the image itself, external to the observer. Once this externalization has taken place, then the decision as to which one to reach for, or which one to avoid, is (from the patient's point of view) made for him.

Real-Unreal Distinction Invariably Accompanies Double Vision in Strabismus

The procedures for demonstrating the presence or absence of nonvisual qualities in the second image constitute an adequate test for, and an operational definition of, the real-unreal distinction. In practice, it is seldom necessary to test all modalities; it is sufficient to ask the patient which of the two images he would approach or reach for if he wanted to touch the object. If the patient's answer is prompt, unequivocal, and correct, then the presence of the real-unreal distinction is demonstrated.

The author has used this procedure in examining over 600 strabismic patients. In every one of these patients awareness of double vision was accompanied by the real-unreal distinction.

REAL-UNREAL DISTINCTION IN NONPARALYTIC STRABISMUS

These observations raised an interesting question: was monocular suppression the end point of the strabismic adaptation, or did the second image continue to deteriorate in perceptual quality even after the patient came to suppress it? To choose between these two alternatives, what was needed was some means of arousing at least a temporary awareness of double vision in patients with well-established nonparalytic strabismus.

Apparatus

Awareness of double vision is enhanced by placing a transparent, colored filter over one eye. In fact, the use of a red glass filter to elicit awareness of double vision is a common clinical procedure and is effective with a large proportion (perhaps one third) of nonparalytic strabismic patients. The general principle involved seems to be differentiation of the two images. With this principle in mind, an apparatus was devised in which the two monocular representations of a single object differed in size, in color, in shape, and in temporal characteristics (one seen continuously, the other seen only intermittently).

The test object was a light box. It contained a transilluminated display consisting of a green disc (diameter, 4.5 centimeters) centered in a red square (outside dimension, 5.7 centimeters; line width, .6 centimeter). The spectral transmission properties of the red and green filters were such that each was opaque to light transmitted by the other. The disc and square were separately illuminated (see Figure 5), so that each could be turned on and off independently of the other. A pair of silent push-button switches, connected to the box by means of a cable, enabled the experimenter to actuate the lamps which illuminated the disc and square. As measured by heterochromatic photometry with a Macbeth Illuminometer, the green disc had a brightness of 19 millilamberts, and the red square had a brightness of 17 millilamberts.

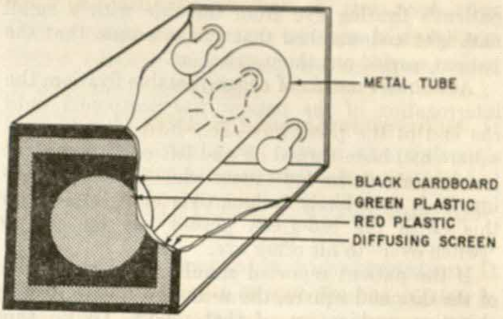


FIG. 5. Cutaway drawing of the disc-and-square apparatus. (The box is about 4 inches square on the end which contains the disc-and-square configuration. The metal tube running the length of the box constitutes a light-tight partition between the disc and square, so that each can be turned on and off independently of the other.)

The patient viewed the light box with a red transparent filter over his deviated eye and a green transparent filter over his nondeviated eye.

The examining room was illuminated by four 100-watt incandescent lamps which were shielded from the patient's direct view. The experimenter controlled the voltage to these lamps by means of a variable autotransformer. The brightness of the surround of the disc and square could be varied continuously from a minimum of .01 millilambert to a maximum of 25 millilamberts.

The standard viewing distance was 6 meters. However, the patient's chair was movable, so that shorter viewing distances were feasible.

Method

The patient was first shown the apparatus and how it worked. He then put on the red-green glasses. The room was darkened (.01 millilambert), the green disc was turned on, and the subject was instructed to look directly at the green disc. The red square was flashed on and off repeatedly 10–12 times, in a cycle of about 1 second on and ½ second off. The patient was then interrogated to determine whether or not he had seen both images at the same time.

The real-unreal distinction is inextricably associated with the pattern of ocular fixation: the real one is the one seen from the fixating eye. For that reason, the patient was repeatedly reminded to continue to "look right at" the green disc. As a further precaution, the examiner illuminated the patient's fixating eye from the side with a small flashlight and watched that eye to ensure that the patient carried out the instruction.

As another means of ensuring stable fixation, the interrogation of the patient was postponed until the end of the procedure—i.e., until after the red square had been turned off and left off. If a patient is asked about the red square while it is still flashing, he is very likely to look over at it; and to do this (with the red-green glasses on) he has to "switch over" to his other eye.

If the patient reported simultaneous perception of the disc and square, the next step was to obtain objective verification of that report. To do this the procedure was repeated, but this time the red square was flashed on and off about 20 times in a random time sequence. The patient was asked to report each time the red square came on. If he was able to do this, he was credited with simultaneous perception of the green disc and red square.

Simultaneous perception having been established, the patient was then interrogated about the appearance of the second image: its color, brightness, position, and other characteristics. In order to answer such questions, the patient frequently asked for another look at the red square. He was given another look, but again every precaution was taken to ensure that he maintained fixation with the eye that saw the green disc. Only in this way could the patient describe the red square as a second image.

Blind-Spot Mechanism

On the retina of every eye, there is a so-called "blind spot" where the optic nerve enters the eye. This spot subtends a visual angle of approximately 6°, centered some 15° to the nasal side of the fovea (Duke-Elder, 1934, p. 896). Thus, a strabismic patient with one eye deviated inward can avoid double vision without having to suppress the second image: he need only maintain an angular deviation of about 15°, and the second image will disappear, so to speak, into the blind spot of the deviated eye. Avoidance of double vision by this means is known as the "blind-spot mechanism" (Swan, 1958).

In the present experiment, whenever a patient with an inward deviation of about 15° failed to see the red square flashing (i.e., whenever the presence of the blind-spot mechanism was suspected), the viewing distance was changed from 6 meters to 2 meters. This had the effect of changing both the size of the second image and (assuming a constant angular deviation) the position of that image on the retina of the deviated eye.

Results

Simultaneous Binocular Perception. Over 600 strabismic patients have so far been examined in this fashion by the present writer. In patients who have been strabismic since birth, prolonged examination (up to 30 minutes) has sometimes been necessary. In patients with extreme amblyopia in the deviated eye, it has sometimes been necessary to reduce the viewing distance to as little as 25 centimeters. However, simultaneous perception of the green disc and the red square, one seen from each eye, was reported and confirmed in every patient.

Apparent Separation of Images. The great majority of strabismic patients reported that the disc and square appeared to be separated in space when both were seen. About 1 patient in 10 described the two images as either concentric or touching, despite the presence of a manifest strabismic deviation; but this "false superimposition" phenomenon was invariably of short duration. After 1 minute of observation every patient had reported that the two images appeared to be separated.

In the present study, no attempt was made to compare the angular separation of images with the angle of ocular deviation.

Presence of the Real-Unreal Distinction. When the two images were seen at the same time and separated, the green disc invariably appeared to be "straight ahead," and the red square appeared off to the side. (The term straight ahead in this context means "correctly localized" or "in the correct place.") The uppermost drawing in Figure 6 shows the disc and square as they might appear to a patient whose left eye is deviated inward.

Patients who were seated more than 1 meter from the light box were asked,

> If you wanted to walk down there and touch the light box, which one would you walk toward, the green disc or the red square?

Patients who were seated within reach of the light box were asked,

> If you wanted to reach out and touch the light box, which one would you reach toward, the green disc or the red square?

Every patient replied that he would walk toward, or reach for, the object as seen from his nondeviated eye—namely, the green disc. These replies demonstrate the presence of the real-unreal distinction in nonparalytic strabismus.

Intensity of the Real-Unreal Distinction. This appearance (green disc straight ahead, red square on the side) is analogous to the visual experience of a patient who had been strabismic for only a year or so (third stage in Figure 4). In both instances, an object (doorknob or light box) is seen double, with one image straight ahead (and real) and the other on the side (and unreal). In nonparalytic strabismus, however, the object as seen from the deviated eye appears much more unreal than anything which is experienced in strabismus of recent onset. The square is seen only momentarily, as it comes on; if it is left on for a few seconds, the patient is likely to report that he no longer sees it. If the room lights are turned up, so that the background of the light box is clearly visible, the patient may no longer be able to see the square, even if it is flashing. At an intermediate brightness level, such that the background is barely discernible, the patient can see the light box, and he may also get an occasional glimpse of the square over on the side. But he does not then see two light boxes. Instead, he sees one box with a green disc in it; and, beside the box he sees a red square which is "hanging in space."

When the square is seen, its position is highly indeterminate; if a small movable green light (seen only from the nondeviated eye) is used as a pointer in an effort to find out exactly where the square is seen, then the patient will report that the square seems to move to avoid the green light. In fact, the patient sometimes describes the square as being continuously in motion "even though it never gets anywhere." The patient has great difficulty in estimating how far the red square is from the green disc and sometimes cannot even tell which side it is on. This last phenomenon, in which the patient sees the second image definitely on the side, but cannot tell which side, was in fact reported by F. H. Verhoeff in 1938 (pp. 670, 675) as an illustration of the fact that the visual experience of strabismic patients cannot be judged by ordinary nonstrabismic standards.

Effect of Covering the Deviated Eye. If the experimenter now covers the patient's right eye, the patient's left eye straightens, as illustrated in Figure 1. At the same time, the red square (as seen by the patient) undergoes a remarkable transformation. It is now easily seen, without flashing, even in ordinary room illumination. Its position is fixed and determinate. As shown in the lower illustration in Figure 6, it now appears (under ordinary room illumination) to be straight ahead, where the green disc was. (The green disc, of course, is no longer seen.) The red square appears to be real and tangible. It is no longer hanging in space, but has a box around it.

All of these things are true whether or not the patient looks directly at the red square. He can look off to the side, so that his eye is in the deviated position; but, so long as his other eye is covered, the square

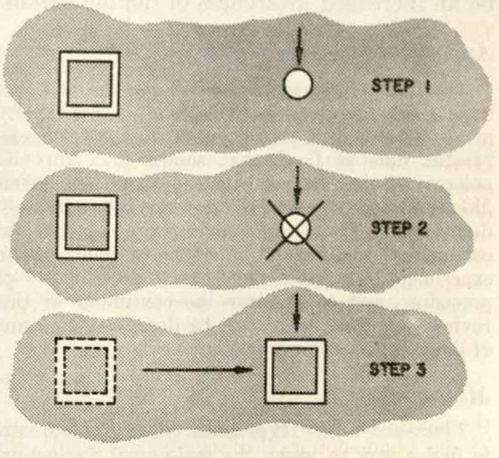


FIG. 6. Green disc (seen only from the right eye) and red square (seen only from the left eye) as they appear to a strabismic patient under appropriate conditions of viewing (see text). (The small vertical arrows—not seen by the patient—indicate the straight ahead position. The patient is assumed to have his right eye straight, left eye deviated inward. In Step 1, the square is perceived as being indeterminate in position, and is not easily seen by the patient. In Steps 2 and 3, the right eye has been covered. The square then comes to appear straight ahead and is easily seen by the patient.)

will retain all of the properties of a real object. The real-unreal distinction is not a matter of eye position, nor is it a matter of what portion of the retina is stimulated. It is a disorder of binocular vision at the level of the visual interpretative mechanism, and it is therefore manifested only when both eyes are open and uncovered.

EXPERIMENT IN AWARENESS OF DOUBLE VISION

The initial observations on paralytic strabismus had pointed to the real-unreal distinction as the mechanism of monocular suppression; and this impression was strengthened by the discovery that the real-unreal distinction is present in intense form in patients who habitually suppress the second image. The next step was to carry out an experimental test of this relationship. If the real-unreal distinction was the basis of strabismic suppression, then lessening of the real-unreal distinction should lessen the patient's tendency to suppress the second image. The result should be an increased awareness of double vision.

Apparatus

The disc-and-square apparatus was used in this experiment. However, in 1957–58, when this experiment (McLaughlin, 1959b) was carried out, the apparatus differed from that shown in Figure 5: polaroid filters were used instead of red and green filters; and the disc and square were not separately illuminated. These differences were matters of convenience and did not affect the rationale of the experiment. To avoid unnecessary descriptions of procedure, and to preserve the continuity of this review, the experiment will be described in terms of the apparatus shown in Figure 5.

Method

Flash-and-Cover Procedure. The problem was to find a way to lessen the real-unreal distinction. It was of no value whatever to have the patient wear a patch over his habitually nondeviated eye. The second image had to be made to take on more and more of the attributes of perceptual reality *while it was still a second image.*

The technique which finally proved successful was the technique which had originally been devised to demonstrate that the second image changes from unreal to real when the nondeviated eye is covered (see Figure 6). The principle involved is this: having seen the second image change from unreal to real, then, when next he sees it over on the side, the patient should perceive it as "a thing that can become real." If the patient can be made to experience this unreal-to-real change repeatedly, the strabismic image should gradually take on more and more of the attributes of perceptual reality.

Another way of describing the rationale of this procedure is that it violates a fundamental operating principle of the perceptual mechanism, namely, the principle that images and objects are disjunct classes. The procedure causes a mere image, a visual artifact, devoid of all nonvisual qualities, to change into a real, substantial object, a part of the environment. Every time this happens, the perceptual mechanism is, so to speak, reminded that the real-unreal distinction is incompatible with the logic of visual perception. By pitting the real-unreal distinction against a more fundamental feature of the perceptual mechanism, the procedure should gradually break down the real-unreal distinction.

Back-and-Forth Exercises. This basic procedure evolved into the "back-and-forth" exercise which is embodied in the instruction sheet shown in Figure 7. The visual exercise described there does not involve covering either eye. That is because, in a dark room and with the red-green glasses on, turning off one image (disc or square) has the same optical effect as covering the eye that sees that image.

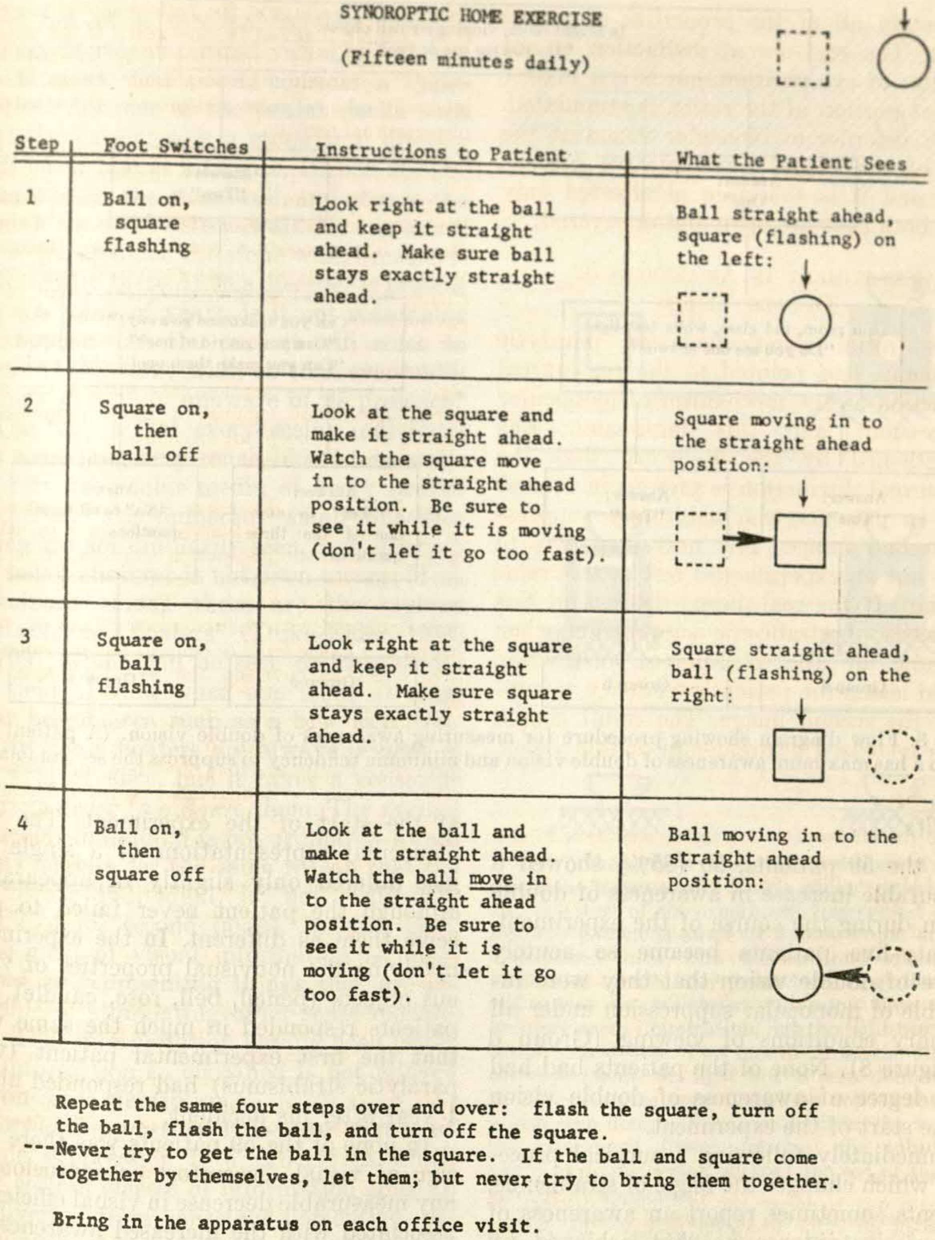
A patient who is practicing the exercise shown in Figure 7 can be seen to move his eyes back and forth repeatedly, so that first one and then the other eye is in the nondeviated position. These eye movements, like the eye movements during cover test (see Figure 1) are more or less concomitant.

Measuring the Awareness of Double Vision. Some strabismic patients cannot be made aware of double vision except with the aid of the disc-and-square apparatus. Some are aware of double vision under ordinary conditions of viewing. Between these two extremes are patients who are aware of double vision only when a red filter is held before one eye, the test object being a small white light in an otherwise darkened room.

These differences among patients make it possible to set up an ordinal scale (Stevens, 1951, p. 25) of awareness of double vision. The flow diagram in Figure 8 shows how such a scale was used in measuring the awareness of double vision in this experiment.

Procedure

To get a measurable lessening of the real-unreal distinction, it proved necessary to have the patients practice the back and forth exercise for 15 minutes daily over a period of several weeks or months. The disc-and-square apparatus was produced in quantity, and one unit was made available to each patient for use in home exercises. Fifty-nine patients, each of whom had a nonparalytic strabismus, participate in the experiment. Each patient visited the laboratory once a week, at which time



SYNOROPTIC HOME EXERCISE

(Fifteen minutes daily)

Step	Foot Switches	Instructions to Patient	What the Patient Sees
1	Ball on, square flashing	Look right at the ball and keep it straight ahead. Make sure ball stays exactly straight ahead.	Ball straight ahead, square (flashing) on the left:
2	Square on, then ball off	Look at the square and make it straight ahead. Watch the square move in to the straight ahead position. Be sure to see it while it is moving (don't let it go too fast).	Square moving in to the straight ahead position:
3	Square on, ball flashing	Look right at the square and keep it straight ahead. Make sure square stays exactly straight ahead.	Square straight ahead, ball (flashing) on the right:
4	Ball on, then square off	Look at the ball and make it straight ahead. Watch the ball move in to the straight ahead position. Be sure to see it while it is moving (don't let it go too fast).	Ball moving in to the straight ahead position:

Repeat the same four steps over and over: flash the square, turn off the ball, flash the ball, and turn off the square.
---Never try to get the ball in the square. If the ball and square come together by themselves, let them; but never try to bring them together.

Bring in the apparatus on each office visit.

FIG. 7. The back-and-forth exercise which has the effect of lessening the real-unreal distinction. (This procedure was first used in experiments in suppression and binocular vision, and is now employed as a therapeutic technique.)

his awareness of double vision was measured by the procedure shown in Figure 8.

Throughout the experiment, the patients practiced no visual exercise except the back-and-forth exercise which is illustrated in Figure 7. However, each patient practiced this exercise under progressively less artificial conditions of viewing (i.e., at progressively higher levels of room illumination) as his awareness of double vision increased. This progression is illustrated in Figure 9.

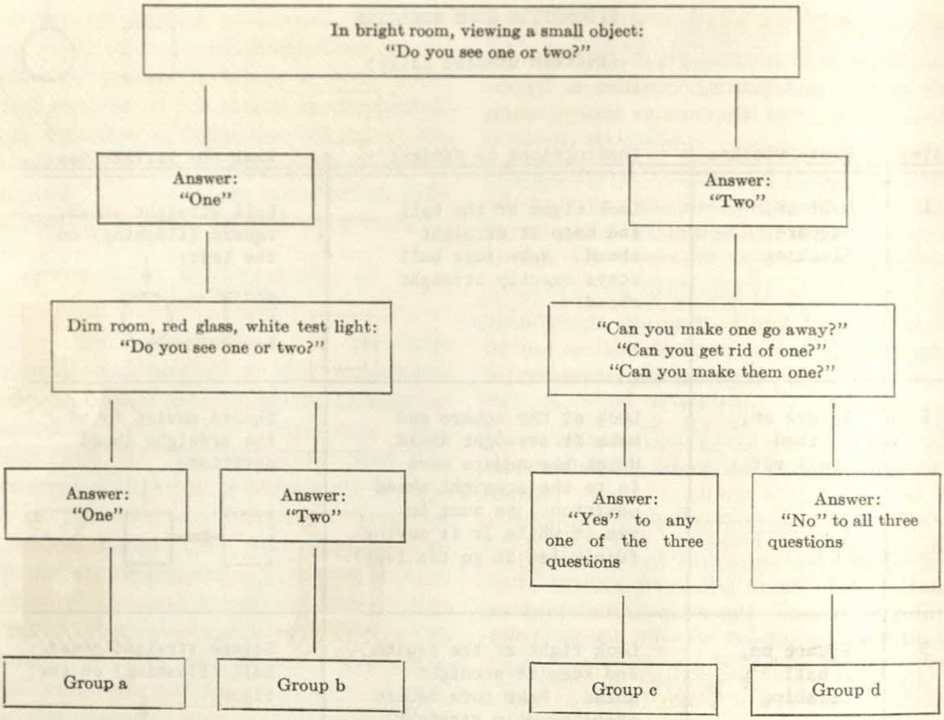

In bright room, viewing a small object:
"Do you see one or two?"
Answer:
"One"
Answer:
"Two"
Dim room, red glass, white test light:
"Do you see one or two?"
"Can you make one go away?"
"Can you get rid of one?"
"Can you make them one?"
Answer:
"One"
Answer:
"Two"
Answer:
"Yes" to any one of the three questions
Answer:
"No" to all three questions
Group a
Group b
Group c
Group d

FIG. 8. Flow diagram showing procedure for measuring awareness of double vision. (A patient in Group a has maximum awareness of double vision and minimum tendency to suppress the second image.)

Results

Of the 59 patients, 50 (85%) showed a measurable increase in awareness of double vision during the course of the experiment. Twenty-five patients became so acutely aware of double vision that they were incapable of monocular suppression under all ordinary conditions of viewing (Group d in Figure 8). None of the patients had had that degree of awareness of double vision at the start of the experiment.

Immediately following a surgical procedure which changes the angle of strabismus, patients sometimes report an awareness of diplopia as intense as that achieved by these 25 patients. But, with that exception, this degree of awareness of double vision does not appear to have been previously reported in the medical literature for patients with concomitant strabismus.

In the patients who became aware of double vision under ordinary conditions of viewing, the real-unreal distinction was still present, though in less intense form than at the start of the experiment. The two monocular representations of a single object differed only slightly in appearance, although the patient never failed to perceive them as different. In the experiment involving the nonvisual properties of various objects (pencil, bell, rose, candle), the patients responded in much the same way that the first experimental patient (with paralytic strabismus) had responded about 1 year after his accident.

In none of the 59 patients was there any sign of visual discomfort or confusion, or any measurable decrease in visual efficiency, associated with the increased awareness of double vision. Evidently, just as in strabismus of recent onset (third drawing from the top in Figure 4), the real-unreal distinction constituted a satisfactory adaptation to the strabismic deviation.

Interpretation of Results

Stone Walls and Floaters. Strabismic suppression appears to represent the abnormal operation of a mechanism that plays a prominent role in normal visual perception. To illustrate this point, consider a hypothetical disorder of the ocular media such that the patient sees a stone wall directly in front of him at all times. He can see objects through the stone wall, and even in the same place as the stone wall. He can never approach or touch the stone wall, because it is not really there. If this disorder is present for a number of years, is it not reasonable to suppose that the patient will come to "ignore" the stone wall, so that eventually he may be totally unaware of its presence?

The fact is that every seeing individual has something very much like "stone-wallitis." In the ocular media of every human eye, there are numerous small opacities which are not ordinarily seen, and of which the naive observer is not even aware. Most prominent among these are the *muscae volitantes* or "floaters" (Duke-Elder, 1934, p. 799), which can be seen drifting across the field of view when one looks at some clear bright area such as a blue sky. Evidently, these floaters are always present in the field of view, but it takes a conscious effort in order to observe them. The normal binocular individual is not ordinarily aware of his floaters for the same reason that the strabismic individual is not ordinarily aware of his second image. In each case, some class of visual information is interpreted as representing things that are incapable of stimulating other sense modalities. That interpretation having been made, the information so classified is not relayed to conscious perception.

Partial Suppression. It is clear that strabismic suppression is a process in which only certain classes of information come to be diverted from conscious perception, the remaining classes being relayed to conscious perception. The fact that a patient can remain well adapted to a strabismic deviation despite having lost the ability to achieve single vision indicates that the perceiving organism is capable of extremely subtle distinctions in choosing the specific classes of information to be suppressed. When pushed to it (as in the present experiment), a strabismic patient can suppress only the bare minimum of information which must be discarded in the interests of visual comfort and efficiency. This property of the adaptive mechanism suggests the possibility of studying the process by which the perceiving organism classifies visual input information.

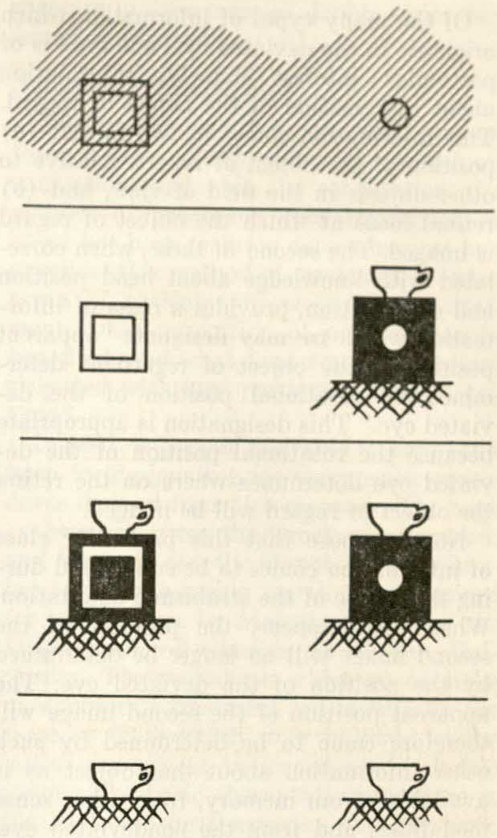


Fig. 9. Progressive change in conditions of viewing in the "awareness of double vision" experiment. (At the start of the experiment [uppermost drawing] the patient is aware of double vision only in a darkened room, with the disc and square apparatus. A month later [second drawing] the disc-and-square are visible at the same time under ordinary room illumination, but the light box itself is not seen double. Two months from the start of the experiment, the light box is seen double, and an object—coffee cup—placed on top of the box is also seen double. Four months from the start of the experiment [lower drawing] the subject is aware of double vision without the aid of the disc and square apparatus.)

(The rate of lessening of the real-unreal distinction which is illustrated here is that which might be anticipated in a 12-year-old patient who had been strabismic since early infancy.)

Of the many types of information which originate in the deviated eye, one type is of particular interest—namely, information about the position of the object of regard. This information comes in two forms: (*a*) position of the object of regard relative to other objects in the field of view, and (*b*) retinal locus at which the object of regard is imaged. The second of these, when correlated with knowledge about head position and eye position, provides a class of information which we may designate "apparent position of the object of regard as determined by rotational position of the deviated eye." This designation is appropriate because the rotational position of the deviated eye determines where on the retina the object of regard will be imaged.

Now, suppose that this particular class of information comes to be suppressed during the course of the strabismic adaptation. When this happens, the position of the second image will no longer be determined by the position of the deviated eye. The apparent position of the second image will therefore come to be determined by such other information about that object as is available—from memory, from other sense modalities, and from the nondeviated eye. All of these other factors work in the same direction: they cause the second image to be relocalized so that it coincides with the primary image.

This line of reasoning clarifies the relationship between (*a*) suppression and (*b*) abnormal retinal correspondence or anomalous projection. Anomalous projection is a phenomenon of partial suppression in which the class "apparent position of object as determined by rotational position of deviated eye" is suppressed, other classes of information from that eye being relayed to conscious perception. A strabismic patient who has single vision with two areas of clear vision (see Figure 3) is evidently suppressing that one specific class of information and is suppressing little else.

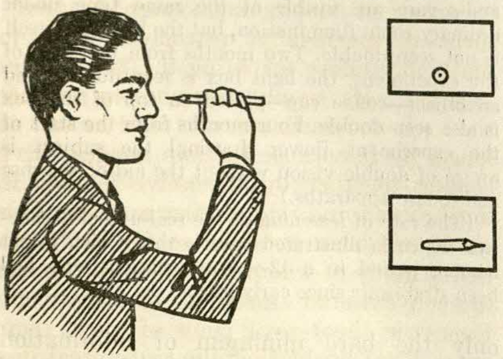


Fig. 10. A viewing situation in which the two monocular representations of a single object differ in size, shape, and color. (The pencil is aligned with the right eye. In the two cut-out drawings at the right, the upper drawing shows the pencil as seen from the right eye, and the lower drawing shows the pencil as seen from the left eye.)

Real-Unreal Distinction as an Obstacle to Binocular Coordination

The 59 patients who participated in the experiment on awareness of double vision were patients who had failed to respond to conventional methods of treatment. During the course of the experiment 7 of them became able, at least temporarily, to hold their eyes in alignment. This was a gratifying result, but of equal theoretical interest was the fact that 19 patients remained strabismic even though they were incapable of monocular suppression. Evidently, the residual real-unreal distinction was enough to cause these patients to remain strabismic.

Perceptual Basis of Normal Binocular Coordination

A normal binocular individual can become aware of double vision (see Figure 2), but such awareness can be maintained only by conscious effort. So long as one is looking intently at some object, the two images seem to be drawn toward one another as if they were magnets having opposite poles adjacent.

The classical and accepted view is that the two images come together in this fashion because of their similarities in shape, color, and size. This rule is generally regarded as having some latitude, in order to take account of situations such as those illustrated in Figure 10 in which two very different images are combined in binocular vision.

The possibility that the real-unreal distinction can be an obstacle to binocular coordination throws new light on this matter. Evidently, in normal binocular vision, the two images are drawn toward one another, not because they resemble each other in visual characteristics, but because they are both interpreted as having the same nonvisual qualities. When one looks at some object—say, a clock on the wall—one has, in a sense, two images of it: one from the right eye and another from the left eye. But each of these is perceived as having the same potentiality for stimulating other sense modalities. One would reach exactly *there* to touch the clock as seen from the right eye, and one would reach toward exactly the same place to touch the clock as seen from the left eye. Because the two are interpreted in this fashion, the appropriate oculomotor reflexes are called into play, the eyes are positioned so as to superimpose the two images, and the two are seen as one. An observer would hardly be credited with normal binocular vision if he were to superimpose two images which did *not* appear to him to represent the same object.

Implications for the Treatment of Strabismus

It is this nonvisual identity between images that is missing in nonparalytic strabismus, even when the patient is aware of double vision. When a strabismic individual sees two separate monocular images of a single clock, only one is interpreted as having any nonvisual qualities, and so the perceptual basis of the normal "rejoin reflex" is absent.

This line of reasoning suggests that, if a patient has a good awareness of double vision, but remains strabismic, then the appropriate treatment is further lessening of the real-unreal distinction. In other words, the back-and-forth exercise should not only bring about an increase in the awareness of double vision; if the patient continues to practice it after having become aware of double vision, it should eventually enable him to hold his eyes in alignment. From the point of view of benefiting the patient, the establishment of binocular coordination is the important result, and the increase in awareness of double vision is a trivial side effect.

Synoroptic Treatment

During the past 5 years, with support from various ophthalmological research organizations and with the cooperation of a number of practicing opthalmologists, an effort has been made to determine just how practical this method of treatment is: what types of patients are amenable to treatment, what proportion of patients can be benefited, and what final result can be anticipated with any particular type of patient?

In 1961, it became desirable to have some term for designating the therapeutic procedures derived from this research. The term *synoroptics* (from the Greek *syn*, together, and *horan*, to see; the branch of optics having to do with seeing together) was coined for this purpose by Malcolm Agnew of the Department of Classics at Boston University. The procedures which go to make up synoroptic treatment are essentially the same as those which were followed in the initial experiment in awareness of double vision. The basic therapeutic procedure is the back-and-forth exercise, and the basic apparatus is the disc-and-square device. A patient under treatment visits the office or clinic at monthly or semimonthly intervals and is expected to devote 15 minutes per day to the performance of home exercises.

A patient undergoing synoroptic treatment is specifically instructed never to make any effort to "bring the two images together." He simply continues to practice the back-and-forth exercise until he can no longer do it because the two images "keep coming together by themselves"—that is, until the normal rejoin reflex appears. This result has been obtained so far in 35 patients, many of whom had failed to respond to conventional methods of treatment. From the theoretical point of view, there seems little doubt that the real-unreal distinction can be, and frequently is, an obstacle to binocular coordination.

It is equally certain that this type of treatment is seldom, if ever, of value by itself. There is little point in lessening the real-unreal distinction in a patient who has, for example, a large uncorrected refractive error or a genuine eye-muscle defect. The method has been effective only when treatment has been carried out in close cooperation with a competent physician.

The proportion of patients who can be treated successfully in this way varies between 5 and 60%, depending on the criteria used in selecting patients for treatment. While these figures serve as a practical guide, from the theoretical point of view they are conservative, for they include as failures those patients who do not cooperate in the treatment or who drop out before having completed a course of treatment. A terminal result usually takes at least 1 year, but some types of patients respond more quickly than that.

It seems likely that every patient who has concomitant strabismus also has the real-unreal distinction. If the real-unreal distinction is an obstacle to binocular coordination, then it should be impossible to cure a strabismic patient unless the real-unreal distinction is lessened. How, then, can we account for the fact that strabismic patients have been cured by conventional procedures? The answer seems to be that some of the standard methods do lessen the real-unreal distinction. For example, surgical modification of the eye muscles, in a child under 6 years of age, will frequently force the child to abandon the real-unreal distinction simply by forcing a change in the angle of strabismus. To return to the analogy of the Stratton experiments, a child under 6 would presumably adapt to reversing lenses more easily than an older patient; and this facility in sensory-motor adaptation makes it possible for him to shift more easily from the strabismic to the nonstrabismic pattern of binocular perception.

THEORY AND TREATMENT OF AMBLYOPIA EX ANOPSIA

Etiology of Suppression Amblyopia

As already noted, it is widely believed that suppression amblyopia derives from monocular suppression. The findings reported in the preceding sections of this paper suggest that this hypothesis might be reformulated to state that suppression amblyopia derives from the real-unreal distinction.

Three lines of evidence support this view. In order to understand the first of these, consider two patients. The first patient has amblyopia ex anopsia in his left eye. His right eye is covered, and he is looking at some object—say, a coffee cup. The second patient has equal vision in the two eyes, but his left eye is deviated inward. He is looking at the same test object (coffee cup), and he sees it double. Each patient is asked to describe the appearance of the coffee cup as seen from his left eye. The remarkable thing is that both patients will use the same terms in describing the coffee cup. To each of these patients, one with amblyopia, one without, the left eye image of the coffee cup appears to be "too small" or "too large"—i.e., indeterminate in size. To both patients, it appears to be "unsteady," continually in motion—i.e., indeterminate in position. Evidently, the perceptual quality of the second image in strabismus without amblyopia is very similar to the perceptual quality of an amblyopic image.

The second line of evidence was obtained by examining patients who had been diagnosed as having amblyopia without strabismus. These patients, when examined with the disc and square apparatus, show the same tendency to suppress the red square (seen from the amblyopic eye) as that shown by a strabismic patient (no amblyopia) when the red square is seen from his deviated eye. Moreover, the red square, when seen, is perceived by the amblyopic observer as being unreal in the sense in which that term is used in the present paper.

The third line of evidence in support of this view is that suppression amblyopia can be treated by lessening of the real-unreal distinction. These results will be presented in Table 1.

Treatment of Eccentric Fixation

Patients with eccentric fixation, like other strabismic patients, possess the innate structure of normal fovea-to-fovea retinal correspondence. This is indicated by the fact that these patients can be made aware of double vision—for example, by means of the disc-and-square apparatus. When examined by the procedure described earlier for arousing awareness of double vision, a patient with eccentric fixation responds in much the same way as any strabismic patient: the green disc (seen from his nondeviated eye) appears to him to be straight ahead, and the red square (seen from his deviated eye) appears to him to be off to the side.

As soon as the patient has reported this appearance, the therapist has at his disposal a powerful technique for treating the eccentric fixation, because the position of the red square over on the side represents, in a sense, the abnormal position of the eye that fixates eccentrically. The therapeutic procedure at this point is to teach the patient to see the red square come in to the straight ahead position when his nondeviated eye is covered (the familiar flash-and-cover procedure).

When he first tries this procedure, the patient is able to see the red square become straight ahead, but his amblyopic eye continues to fixate eccentrically. However, as he continues to practice it, the eye can be seen to make a rudimentary fixational effort—i.e., to move part of the way toward foveal fixation—each time the red square comes in to the center. After a few weeks or months with daily practice of this procedure, the patient will take up foveal fixation with his amblyopic eye each time the red square comes in to the center. By progressing to less and less artificial conditions of viewing, the patient can become able to fixate foveally with that eye under ordinary conditions of viewing. This method has been used successfully with adults having gross eccentric fixation.

It will be understood that the foregoing account of clinical procedure is intended to convey the rationale of the method, and is not intended to equip the reader to treat eccentric fixation.

Treatment of Amblyopia ex Anopsia

The most common form of treatment for amblyopia ex anopsia is to have the patient wear an opaque occluder over his unaffected eye, thus forcing him to use his amblyopic eye. While this method is often valuable and is sometimes the only feasible procedure, it has certain drawbacks. Wearing an eye patch may be a considerable burden to the patient, both in terms of visual handicap and in terms of cosmetic defect. In young patients, prolonged wearing of the patch may simply reverse the pattern of ocular fixation, so that the eye behind the cover becomes the one that is amblyopic. If there is a latent deviation, prolonged occlusion of one eye may cause the deviation to become manifest.

The other widely used type of treatment for amblyopia ex anopsia consists of active stimulation of the amblyopic eye, with particular emphasis on the use of brightly colored, high contrast material having intrinsic interest to the patient (Kramer, 1953, p. 360). The unaffected eye is occluded during the period of active stimulation, but may be left uncovered at other times. This type of treatment has recently received considerable impetus from the work of Bangerter (1955), who uses it in conjunction with a special technique of localized retinal stimulation as a means of eliciting foveal fixation.

The visual exercise illustrated in Figure 8 requires that the patient repeatedly achieve accurate fixation and accurate localization with his amblyopic eye, and that he accomplish these things with both eyes open and uncovered. At one stage of the exercise (Step 3), the patient is required to maintain accurate fixation with his amblyopic eye even while the other eye is being stimulated. These features of the exercise suggested that it might provide an effective mode of treatment for amblyopia ex anopsia.

Table 1 shows some results of this type of treatment. Thirty-one patients participated in the experiment, and 15 showed a measurable improvement in visual acuity during the course of the experiment. In these results no effort has been made to separate changes in centricity of fixation from changes in suppression amblyopia. The results shown in Table 1 are compara-

TABLE 1

RESULTS IN THE TREATMENT OF AMBLYOPIA EX ANOPSIA BY THE PROCEDURE DESCRIBED IN THE TEXT

Patient	Age	Column A		Column B	
		Date	Vision	Date	Vision
BA	5	8/25/61	5/200	1/22/62	20/100
MD	11	8/4/61	20/70	1/8/62	20/50
AM	10	9/8/61	20/70	1/22/62	20/50
AW	10	8/31/61	20/200	11/22/61	20/70
RB	39	9/22/61	20/70	9/30/61	20/20
PC	17	9/8/61	20/100	1/10/62	20/50
BD	8	10/6/61	20/200	1/9/62	20/50
LH	6	10/19/61	20/200	1/11/62	20/50-1
RL	19	7/18/61	20/70	8/9/61	20/30
JM	6	9/20/61	20/200	12/11/61	20/70
HM	16	9/14/61	20/200	1/26/62	20/40-1
DN	7	9/29/61	20/200	1/16/62	20/40-2
MN	11	7/24/61	20/200	1/11/62	20/100
JV	30	11/15/61	20/40	1/17/62	20/30
LB	7	7/24/61	20/200	1/26/62	20/50

Note.—Column A shows the date treatment was started and the visual acuity in the amblyopic eye at that time. Column B shows the date at which the second measure of visual acuity was made (not necessarily the date of termination of treatment). Visual acuity measures were made with Snellen test letters at 20 feet. The 15 patients reported in this table represent the successful results in a group of 31 patients; the remaining 16 patients failed to respond to treatment.

ble to results which have been reported by other investigators for patching and active stimulation (Priestley, 1963, pp. 621–622; Schlossman, 1961, pp. 844–845).

The most outstanding feature of synoroptics as a mode of treatment for amblyopia ex anopsia is its practicability. Patients who participated in the experiment described here devoted a maximum of 15 minutes per day to home exercises, visited the laboratory only once a month, and did not wear a patch over either eye at any time during the course of the experiment. In contrast to this, the treatment of amblyopia ex anopsia by conventional means may require supervised treatment in the office or clinic for as much as 10 hours per week (Romanio, 1962, p. 7).

OBSERVATION OF THE REAL-UNREAL DISTINCTION

The real-unreal distinction is so obvious a feature of nonparalytic strabismus, and so enhances our understanding of this complex disorder, that some explanation is in order as to why it was not described in the medical or scientific literature until 1959 (McLaughlin, 1959b, p. 86). An account of the factors which caused the real-unreal distinction to escape detection for so long may also aid other investigators in observing and studying this phenomenon.

Use of Real Objects as Test Stimuli

During the past century, almost all research in strabismus has been done with optical devices, such as the stereoscope, which make it easy to present controlled stimuli separately to each eye. Visual stimuli presented by such devices have nonvisual qualities which are at variance with their properties as visual stimuli. In a typical viewing situation, the patient looks into an elaborate binocular instrument and sees a "lion" from his right eye and a "cage" from his left eye. (The lion and the cage are drawings on lantern slides which have been inserted into the instrument.) The patient is then asked such questions as, "Is the lion in the cage?" The patient knows quite well that, instead of a lion and a cage, there is a man in a white coat in front of him, whom he could see if the machine were not in the way. Both the lion and the cage are unreal. Neither one can be straight ahead, and it makes no sense to ask the patient which one he would approach if he wanted to touch some object

It is impossible in this situation even to detect the real-unreal distinction, much less to study it or to attempt to lessen it.

The disc-and-square apparatus—especially when viewed in a dark room—provides a viewing situation which is in some respects more artificial than the lion-and-cage display. However, the disc-and-square apparatus is a real object. If the patient has any doubt about that fact, he can walk across the room and find out for himself: it can be touched and manipulated, and it makes a noise when he taps on it. It is where it appears to be, and it is what it appears to be.

The cardinal requirement for detection of the real-unreal distinction is that the test stimulus be perceived by the patient as having nonvisual qualities which are not at variance with its properties as a visual stimulus.

Patient not Consciously Aware of the Distinction

Another reason why the real-unreal distinction is not well known is that the patient himself is not consciously aware of it and may deny its existence. It is not uncommon for an intelligent and cooperative patient to report that both images appear to him to be equally real. But if the object is, for example, a knife, and if the patient is asked, "Which one would you use to cut with?" then he will choose the knife as seen from his nondeviated eye. Or, if the object is a coffee cup, and he is asked, "Which one would you pour coffee into?" his answer will again reveal the presence of the real-unreal distinction. For these demonstrations, it is desirable to use a red filter over one eye so as to enable the patient to name the two images separately.

The simplest demonstration of all is this: if the patient has strabismus, and if he can find his way around unaided, with both eyes open and uncovered, then he has the real-unreal distinction.

Never Present in Normal Binocular Vision

There is still another reason why the real-unreal distinction has escaped detection for so long. A normal binocular individual, when he is aware of double vision, may believe that he experiences the phenomenon himself. In fact, he has only to decide arbitrarily that one image is the real one; and the other image then looks unreal. The conclusion is obvious: if the real-unreal distinction is this easy to experience, it can hardly be a serious and often insurmountable obstacle to binocular coordination.

The fact is that a person with normal binocular vision cannot possibly experience the real-unreal distinction as it exists in nonparalytic strabismus. The strabismic real-unreal distinction represents the accumulation of many years of sensory-motor experience, usually including the period of infancy when sensory-motor maturation is the chief business of living. It constitutes an array of visual skills which a nonstrabismic individual cannot imitate, and a visual experience which a nonstrabismic individual cannot have.

STRABISMUS AND PERCEPTUAL RESEARCH

Strabismus may prove to be a valuable research tool for the study of certain phenomena which are not otherwise accessible to experimental investigation. The fundamental psychological process of repression surely has a close analogy in strabismic suppression, and the latter is readily amenable to systematic experimentation. Strabismus is a unique "sensory-deprivation" experiment, in which only one eye is deprived of its intersensory correlations, and in which a period of deprivation of 40 years is not uncommon. Strabismus appears to be the only disorder of visual perception in which the patient's reliability as a scientific observer, and his ability to report his visual experience, are unimpaired.

Strabismic individuals report many phenomena which throw light on nonstrabismic perceptual mechanisms. For example, during the procedure illustrated in Figure 6, when the patient is instructed to see the square move in to the straight ahead position, he frequently rejects the unreal-to-real change by simply failing to see the square move. He sees the square undergo a gross change in position; and he sees it continuously as it does so; but no movement is perceived. (In fact, it has been necessary to devise a battery of clinical techniques for "forcing" perception of movement during this procedure.) The fact that the patient can reject the perception of movement in this fashion confirms what Wertheimer pointed out in 1912 (see Boring, 1942, p. 595)—namely, that the perception of movement is a unique and unitary phenomenon not fully defined or determined by the conditions which ordinarily surround it.

Most intriguing of all, though, is the insight which strabismus affords into visual perception as an organizational process. In the strabismic patient's detailed description of the red square "hanging in space," we see an intact, mature sensory system providing grossly disorganized sensory data directly to conscious perception, without benefit of the data-processing procedures which ordinarily classify and regroup the data so as to create an external environment consisting of objects in space. Then, as the perceptual disorder is lessened, we see the organizational process gradually intervene. What makes strabismus so interesting is that it affords a rare glimpse, and the hope of a better view, of this critical area of interaction between organism and environment.

SUMMARY

Strabismus is a visual disorder in which one eye is deviated from the line of sight. Amblyopia (more properly, *amblyopia ex anopsia*, which means deterioration of vision through disuse) is a partial deficiency of visual acuity in one eye—usually in a strabismic eye. Evidence is presented that the underlying disorder in the strabismus-amblyopia syndrome is a previously unrecognized anomaly of visual perception: objects as seen from one eye (the strabismic or amblyopic eye) are perceived as being unreal. The term *unreal* in this context means incapable of stimulating any sense modality except the visual. Operations appropriate to this definition are described. Since objects as seen from the unaffected eye are perceived in the normal fashion, the patient may be said to make a "real-unreal distinction" between the two monocular representations of the object of regard.

The real-unreal distinction was first studied in a patient who had recently acquired paralytic strabismus as the result of an injury to his extraocular musculature. As the patient adapted to his strabismic condition (over a period of about 2 years), the unreal image underwent a progressive change in appearance. There was no change in the eye itself which might have accounted for this change in appearance. The end result of this adaptive process appeared to be monocular *suppression*—i.e., the disappearance of the unreal image from conscious perception.

In adults who have been strabismic from infancy, monocular suppression is intense: such patients cannot be made aware of the unreal image except under highly artificial conditions of viewing. When a patient of this type is made aware of the unreal image, his verbal responses indicate the presence of the real-unreal distinction in a much more intense form than that observed in strabismus of comparatively recent onset.

These observations suggested that the real-unreal distinction might be the mechanism of monocular suppression in strabismus. To test that hypothesis, a technique was devised for lessening the real-unreal distinction. The technique was applied to 59 strabismic patients, 50 of whom (85%) acquired a measurable increase in awareness of the unreal image. It was concluded that the hypothesis was correct: strabismic suppression is a process in which the image becomes, in the words of one patient, "so unreal that I don't even see it."

Of the 59 strabismic patients who participated in the experiment, 7 became able, at least temporarily, to hold their eyes in alignment. This finding suggested that the real-unreal distinction might actually be an obstacle to binocular vision and might account for the persistence of the strabismic condition in some patients. This proved to be the case. Lessening of the real-unreal distinction has proven to be, in many patients, an effective adjunct to the use of surgery and eyeglasses in the treatment of strabismus.

Strabismus is, in effect, a sensory-deprivation experiment in which a period of deprivation of 40 years is not uncommon, and in which the subject's reliability as an observer is nevertheless unimpaired. When a strabismic patient describes the characteristics of the unreal image, he is describing something which has no counterpart in normal visual perception. He is giving an account of the operation of a visual mechanism in which raw sensory data are relayed directly to conscious perception without benefit of the data-processing procedures which ordinarily intervene to produce an orderly visual environment consisting of objects in space. On subsequent examinations over a period of several months, as his perceptual disorder is gradually alleviated, he can describe the successive stages of approach to normal visual perception.

REFERENCES

BANGERTER, A. *Amblyopiebehandlung*. Basel: S. Karger, 1955.

BORING, E. G. *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century, 1942.

DUKE-ELDER, W. S. *Textbook of opthalmology*. Vol. 1. St. Louis: C. V. Mosby, 1934.

HEBB, D. O. Sensory deprivation: Facts in search of a theory. Discussion. *J. nerv. ment. Dis.*, 1961, **132**, 40–43.

HELMHOLTZ, H. *Physiological optics*. Vol. 3. (3rd ed. trans. by J. P. C. Southall) New York: American Institute of Physics, 1925.

HENDERSON, J. W. The neurology of amblyopia. In G. M. Haik (Ed.), *Strabismus: Symposium of the New Orleans Academy of Ophthalmology*. St. Louis: C. V. Mosby, 1962.

JAMES, W. *The principles of psychology*. Vol. 2. (Orig. publ. 1890) New York: Dover, 1950.

KEINER, G. B. J. *New viewpoints on the origin of squint: A clinical and statistical study on its nature, cause, and therapy*. Hague: Nijhoff, 1951.

KRAMER, M. E. *Clinical orthoptics*. St. Louis: C. V. Mosby, 1953.

LINKZ, A. *Vision: Physiology of the eye*. Vol. 2. New York: Grune & Stratton, 1952.

MCLAUGHLIN, S. C. The elicitation of fusion in strabismus patients who are aware of diplopia. *Amer. J. Ophthal.*, 1959, **48**, 148–153. (a)

MCLAUGHLIN, S. C. Increasing the awareness of diplopia in strabismic patients. *Amer. Orthoptic J.*, 1959, **9**, 77–88. (b)

MCLAUGHLIN, S. C. An hypothesis concerning the etiology of nonparalytic strabismus. *Amer. J. Ophthal.*, 1960, **49**, 1218–1221.

MCLAUGHLIN, S. C. Synoroptics. *Amer. Orthoptic J.*, 1964, **14**, 134–143.

NEWTON, I. *Opticks*. (Based on the 4th ed., 1730). New York: Dover, 1952.

PRIESTLEY, B. S., HERMANN, J. S. & NUTTER, A. H. Home pleoptics. *Arch. Ophthal., Chicago*, 1963, **70**, 616–624.

ROMANIO, J. A. A report on clinical experience in pleoptics. *Amer. Orthoptic J.*, 1962, **12**, 7–11.

SCHLOSSMAN, A. *Prognosis, management, and results of pleoptic treatment*. Vol. I, Boston: Little, Brown, 1961.

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, New York: Wiley, 1951.

STRATTON, G. M. Some preliminary experiments in vision without inversion of the retinal image. *Psychol. Rev.*, 1896, **3**, 611–617.

SWAN, K. C. The blind spot mechanism in strabismus. In J. H. Allen (Ed.), *Strabismus Ophthalmic Symposium* II. St. Louis: C. V. Mosby, 1958.

TEUBER, H.-L. Sensory deprivation: Facts in search of a theory. Sensory deprivation, sensory suppression and agnosia: Notes for a neurologic theory. *J. nerv. ment. Dis.*, 1961, **132**, 32–40.

VERHOEFF, F. H. Anomalous projection and other visual phenomena associated with strabismus. *Arch. Ophthal., Chicago*, 1938, **19**, 663–699.

(Received April 28, 1964)

Vol. 78, No. 13 Whole No. 590, 1964

Psychological Monographs: General and Applied

IMPROVING THE ACADEMIC PERFORMANCE OF ANXIOUS COLLEGE FRESHMEN:

A GROUP-COUNSELING APPROACH TO THE PREVENTION OF UNDERACHIEVEMENT[1]

CHARLES D. SPIELBERGER and HENRY WEITZ

Vanderbilt University *Duke University*

Anxious college freshmen were invited early in their first semester to participate in a special academic orientation program. Those who volunteered were divided into experimental (E) and control (C) groups. Students in the E group were assigned to counseling groups which met each week of the first semester. The C group was invited to participate in counseling groups during the following semester along with those members of the E group who wished to continue. Students who regularly attended first-semester counseling groups made higher grades than students who either did not regularly attend or who were assigned to the C group. The effects of counseling during the second semester were obscured by the adverse influence of fraternity affiliation on grades. The influence of systematic trends in university admission policies on the relationship between anxiety and academic performance were discussed, as were differences in the relationship between anxiety and grades for anxious students who volunteered to participate in the counseling groups as compared to nonvolunteers.

The observation that underachievement is frequently associated with personality problems has been made in a number of recent investigations of the academic adjustment of college students. For example, Blaine and McArthur (1961) identified a group of underachieving college students with conflicts which caused them to seek academic failure as a means of expressing rebellion against parents toward whom they felt strong, unconscious resentments. Walters (1961) ascribed somewhat similar personality dynamics to underachievers for whom aggressiveness in academic work had become associated with the expression of repressed anger. For these students, apathy, which served as a defense against the expression of hostility, was the salient personality characteristic. It has also been demonstrated that anxious college students earned lower grades (Katzenmeyer, 1958; Spielberger & Katzenmeyer, 1959) and, on long-term follow-up, had a higher dropout rate due to academic failure (Spielberger, 1962) than nonanxious classmates of comparable ability.

There is little evidence, however, that personality problems are direct and immediate causes of poor academic performance. It seems more likely that, in response to the pressures of college life, students with personality problems are predisposed to develop maladaptive study habits and attitudes which, in turn, interfere with the learning process and lead to underachievement. For college students identified as having personality problems, preventive measures implemented at the beginning of the freshman year would come at a time when the potential for serious maladaptive behavior is heightened by new environmental stresses.

General recognition of the importance of the first semester of college for the subsequent academic adjustment of college students is attested to by the number of institutions of higher learning which conduct orientation programs for entering freshmen. Aside from the informational aspects of such programs, however, much valuable professional time and effort may be spent with students who would do relatively well without the benefits of any special intervention. Furthermore, since the determinants of underachievement in many cases are multiple and complex, the associated personality conflicts of underachievers frequently require the attention of professional persons trained in counseling or psychotherapy. Thus, it would seem unlikely that a *general* orientation program would provide an effective means of facilitating the academic adjustment of most potential underachievers.

---

[1] This investigation was supported by a Mental Health Demonstration Project Grant (OM-362) from the National Institute of Mental Health, United States Public Health Service. The authors wish to express their appreciation to Edna Bissette, J. P. Denny, J. B. Grier, Kay Howard, and E. S. Katkin for their assistance in the collection and analysis of the data.

An added difficulty in working with underachievers is that they are apparently reluctant to face their underlying emotional conflicts. Blaine and McArthur (1961) reported that the underachievers with whom they worked tended to

> have as a distinguishing feature an extreme reluctance to seek or accept any help for the basic problems underlying their academic difficulty [p. 89].

These students sought help only when they were on the verge of dropping out of school and then primarily upon direct referral from a dean or faculty member. Similarly, Walters (1961) observed that apathetic underachievers presented themselves at the student mental health service only when in serious academic difficulty or when sent there because of academic deficiencies. Furthermore, where counselors have attempted to work with underachievers, it has been found that although counseling may facilitate the personal adjustment of such students it has little demonstrable effect upon their grades (Baymur & Patterson, 1960; Broedel, Ohlsen, Proff, & Southard, 1960).

If, indeed, there are students with personality problems for which they are reluctant to seek help, and which predispose them to develop maladaptive study habits and attitudes, assisting such students to realize their academic potential involves both finding a way to identify them and determining a means for aiding them to utilize their capabilities more fully. The goals of the present study were to explore the utility of group counseling, offered at the beginning of the freshman year, for preventing or ameliorating the academic adjustment difficulties of anxious college freshmen. On the basis of previous research findings (Katzenmeyer, 1958; Spielberger, 1962; Spielberger & Katzenmeyer, 1959), it was expected that such students would become academic underachievers if left to their own resources. Furthermore, it was assumed that anxious freshmen would be more likely to accept counseling than would potential underachievers without anxiety, since anxiety is generally considered to be associated with positive motivation to seek assistance.

METHOD

Subjects and Research Design

This investigation was carried out with two successive classes of male liberal arts freshmen who entered Duke University in the fall terms of 1959 and 1960. The experimental procedures employed in both years are summarized in Figure 1 and may be divided into six distinct steps. First, anxious freshmen were identified as early as possible in their first semester at the university in order to permit counseling to be initiated prior to the students' experiencing excessive failure or frustration. Second, anxious freshmen who scored below the fourth stanine on local norms of a measure of scholastic aptitude were judged to be less likely to profit from counseling and were, therefore, eliminated from further consideration. Third, the anxious freshmen who met the scholastic aptitude criterion were invited to participate in a special academic orientation program. Fourth, students who volunteered were assigned to experimental and control groups which were matched on variables known to be related to academic performance. Fifth, students in the experimental group were assigned to small groups in which counseling procedures were carried out. Finally, the effects of counseling were evaluated by comparing the academic performance of counseled students, taking into account the extent of their participation in the counseling groups, with appropriate control groups of noncounseled students.

The second year of the study was initially viewed as a replication of the first year. Therefore, it was originally intended that the same experimental procedures would be employed for both the 1959 and 1960 classes (samples). However, although the general approach was similar in the 2 years, the specific experimental procedures unavoidably differed in several important respects. There were also unexpected but important differences in the intellectual characteristics of the two samples. These differences in procedures and sample characteristics are described below.

Identification of Anxious College Freshmen

All freshmen entering the university in the fall semester of 1959 and 1960 were given a modified form[2] of the Minnesota Multiphasic Personality Inventory (MMPI). This modified MMPI included all of the items comprising the three validity scales, the nine original clinical scales, the Taylor (1953) Manifest Anxiety (*MA*) scale, and the Welsh (1956) Anxiety (*A*) scale. A biographical data card, on which students indicated the type of high school they had attended and whether or not they had decided on a major area of study, was completed during both years. In 1959 the American Council on Education (ACE) Psychological Examination, 1949 edition, was also given.

All of the tests were administered as integral parts of the university's regular placement testing program which is conducted during freshman orientation week prior to the beginning of classes. Although all entering freshmen were tested, counseling was offered only to anxious male students since the adverse relationship between anxiety and academic performance had been found only for college men (Spielberger, 1962). The study was further restricted to students enrolled in the liberal arts curriculum in order to insure that the students who participated would be taking comparable courses throughout their freshman year.

Identification of Anxious Freshmen in 1959. The criteria for the identification of anxious freshmen in 1959 required that a student score in the upper 30% of the distribution of scores on the *MA* scale (16 and above) and in the upper 50% of the *A*-scale score distribution (10 and above). In previous studies college men with comparable scores on the *MA* scale had obtained lower grades and had a higher percentage of academic failures than nonanxious students of equal ability. In addition to the *MA* scale, the *A* scale was employed in the present investigation to insure that only students who were consistently high in maladaptive anxiety would be included. Of the students satisfying the anxiety criterion in 1959, those who scored below the fourth stanine on local ACE norms (raw scores of 104 or below) were eliminated since the previous research had indicated that students with limited aptitude showed marginal academic performance irrespective of their anxiety level. Data were available for 552 of the 625 male liberal arts freshmen who entered the university in 1959. The majority of the students on whom data were missing, both in 1959 and in 1960, failed to complete the modified MMPI. Of those students for whom complete data were available in 1959, 471 were at or above the fourth stanine on the ACE, and 112 met the combined anxiety and intellectual criteria.

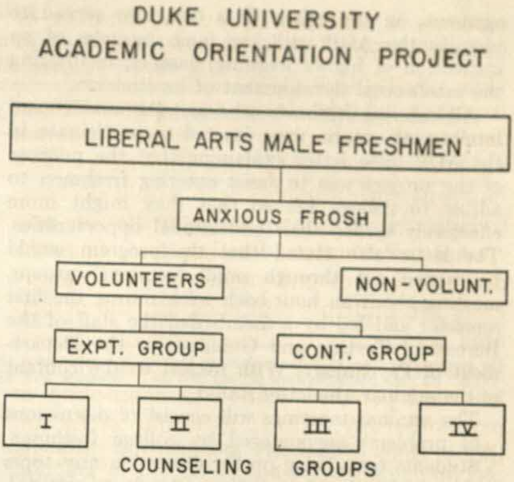


FIG. 1. The experimental design of the Duke University Academic Orientation Project.

Identification of Anxious Freshmen in 1960. The criteria for identifying anxious freshmen to be offered counseling in 1960 were similar to those employed in 1959. The *MA*-scale and *A*-scale cutoffs were established at raw scores of 16 and 12, respectively. ACE scores were not available in 1960 due to the discontinuation of the routine administration of this test with the advent of a university requirement that all applicants for admission take the College Entrance Examination Board (CEEB). Therefore, for the 1960 sample it was necessary to establish a CEEB cutoff score which was comparable to the 1959 ACE cutoff score. It was determined that students who met the anxiety criteria and who scored at or above the third stanine on university CEEB norms (V plus M = 929) would be invited to participate in the counseling groups. Complete data were available for 490 of the 572 male liberal arts freshmen who entered the university in 1960; of the 434 students who met the CEEB criterion, 122 also met the anxiety criteria.

Inviting Anxious Freshmen to Participate in the Academic Orientation Project

The total program of research was consistently referred to as the Academic Orientation Project (AOP) in an effort to divorce it in the minds of students and faculty from overtones of psychotherapy, on the one hand, and administrative management, on the other. This title also served to identify the AOP with the main business of an institution of higher learning, namely, facilitating the intellectual development of its students.

---

[2] The MMPI was shortened from 566 to 398 items so that it could be completed in the limited time available (1¼ hours). We are grateful to H. Seashore and the Psychological Corporation for granting permission to modify the MMPI for this investigation.

All entering freshmen who met the anxiety and intellectual criteria were invited to participate in the AOP by a letter explaining that the purpose of the project was to assist entering freshmen to adjust to college life so that they might more effectively utilize their educational opportunities. The letter also stated that the program would be carried out through small "seminar" groups, meeting about an hour each week during the first semester and led by a member of the staff of the Bureau of Testing and Guidance or the Department of Psychology. With respect to the content of the seminar, the letter stated,

> The seminar meetings will consist of discussions of problems encountered by college freshmen. Students may bring up for discussion any topic they wish related to their academic, vocational, personal, and social interests. In addition to the group discussions, students who want individual assistance will be helped to find appropriate resources for such aid, for example, those wanting educational or vocational counseling will be referred to the Bureau of Testing and Guidance.

In order to reduce further the administrative and psychotherapeutic overtones of the AOP, the letter emphasized that participation by those invited was entirely voluntary. Students who wished to participate in the AOP were asked to return a completed schedule card and to commit themselves to participation for a minimum of 4 weeks. The latter requirement was instituted because it was expected that mounting pressure from academic work might cause some students to drop out before they realized any of the possible benefits of the program.

It was intended that the initial invitation to participate in the AOP would reach students on or before the first day of classes. Unfortunately, in 1959, almost one third of the students failed to complete the MMPI during the allotted time necessitating the scheduling of an additional testing period. The identification of anxious freshmen was further delayed by unforeseen scoring and coding problems. Consequently, in 1959 the initial invitation was not mailed until 9 days after the beginning of classes.

In 1960 the instructions for the MMPI were modified. Students were prompted to respond more rapidly to the MMPI, which was given as the final test in the placement battery. Students who did not complete it in the allotted time were urged to remain until they had done so. Thus, in 1960 it was possible to mail letters inviting anxious freshmen to participate in the AOP early enough so that the letters were received 1 or 2 days prior to the first day of classes. In both 1959 and 1960 a follow-up letter was sent 6 days after the initial invitation to participate was mailed.

Of the 112 students invited to participate in 1959, 56 (50.0%) volunteered for the project. In 1960 75 of 122 anxious freshmen (61.5%) volunteered. Although a larger percentage of invited students volunteered for the AOP in 1960, the difference in the proportion of volunteers between the 2 years was not statistically significant ($\chi^2 = 2.67$, $df = 1$). The pattern of volunteering, however, was markedly different for the 2 years. Of those invited in 1959, only 26 students (23.2%) volunteered in response to the initial letter; in 1960 66 students (49.2%) responded to the initial letter. This difference in response to the initial letter was highly significant ($\chi^2 = 24.55$, $df = 1$, $p < .001$) and was probably due in part to the fact that the initial invitation to participate reached the anxious students earlier in 1960 than in 1959 and possibly at a time when their anxiety about subsequent academic adjustment was relatively intense. It also seems likely that favorable publicity about the AOP in the campus newspaper at about the time that students received their invitations to participate influenced early volunteering in 1960.

Assignment of Volunteers to Counseling and Control Groups

In both 1959 and 1960 anxious freshmen who volunteered for the AOP were divided into an experimental (E) group, which was to be counseled, and a noncounseled control (C) group. In 1959 the E and C groups were carefully matched on factors found previously to relate to academic performance at the university. These included ACE scores (Spielberger & Katzenmeyer, 1959), type of high school attended (Weitz & Wilkinson, 1957), and the declaration of a major field of study (Weitz, Clarke, & Jones, 1955). The latter two variables, however, were not found to be related to academic performance for the 1959 sample. Therefore, in 1960 the E and C groups were matched only on scholastic aptitude. The measure of scholastic aptitude employed in 1960 was CEEB (V plus M) score. The means and standard deviations, for the measures on which the E and C groups were matched in 1959, are presented in Table 1 for both samples. There were no statistically significant differences between the E and C groups in either sample on any of these measures.

Each student in the E group was notified by letter of the time and place that the counseling group to which he had been assigned would meet. In both years students were assigned to counseling groups on the basis of the periods of free time which each indicated, when he initially volunteered for the AOP, that he would have available. There were four counseling groups in 1959 and three in 1960. The counseling groups were nearly equal in size in 1959, with 6–8 students assigned to each group, but this was not possible to arrange in 1960. The three groups numbered 8, 11, and 15 students in 1960.

Each student in the C group was notified by

TABLE 1

Means and Standard Deviations on Matching Variables* for Anxious Volunteers Assigned to Counseled and Noncounseled Groups

Group	*N*	ACE/CEEB		% private school		% major declared	
		M	*SD*	*M*	*SD*	*M*	*SD*
1959							
Counseled (E)	26	130.8	16.2	15.4	7.0	19.2	8.4
Noncounseled (C)	27	132.2	15.1	11.1	6.0	25.9	8.4
1960							
Counseled (E)	34	1206.3	113.8	14.7	6.1	29.4	7.8
Noncounseled (C)	38	1214.3	112.4	34.2	7.7	15.8	5.9

* The 1959 groups were matched on ACE, percentage of students who attended private secondary schools, and percentage of students who declared a major area of study at the beginning of the first semester. The 1960 groups were matched only on CEEB scores.

letter that more students had volunteered for the AOP than could be worked with during the fall semester, and that he would be offered an opportunity to participate during the second semester. These students were also advised of the availability of the resources of the University Bureau of Testing and Guidance in the event that they desired to discuss academic, vocational, or personal problems with a qualified counselor prior to the second semester.

Group Counseling Procedures: First Semester

Group counseling was conducted within the administrative framework of the University Bureau of Testing and Guidance. Although it was planned that the counseling sessions would begin shortly after classes commenced, it was not possible in 1959 to schedule the first group-counseling session until the end of the fifth week of classes because of the unforeseen administrative difficulties which delayed sending the invitations to anxious freshmen. In 1960, however, the initial meetings with counseling groups were held during the second week of classes as originally planned.

The number of counseling sessions conducted in 1959 ranged from 8-11. One group terminated prior to the Christmas vacation; the other three groups met until the end of the semester. The 1960 groups met for either 12 or 13 sessions, and all three groups continued until shortly before the end of the semester. The average number of sessions attended by the participants in 1959 was 8.7 with a range of from 2 to 11 sessions. In 1960 the average number of sessions attended was 7.00 with a range of from 1 to 12 sessions. Although students had been asked to commit themselves to attend at least 4 sessions, no effort was made to enforce this requirement.

The two counselors were members of the university faculty; both had extensive previous experience in group and individual counseling and/or psychotherapy with college students. One of the counselors was an American Board of Examiners in Professional Psychology (ABEPP) diplomate in counseling psychology prior to the beginning of the AOP; the second achieved the ABEPP diploma in clinical psychology during the second year of the project. The counselors had both been at the university for at least 4 years when the project was initiated; hence, they were familiar with university requirements, regulations, and folkways. Each counselor worked with two groups in 1959; the counselor trained in counseling psychology worked with two of the three groups in 1960. In both years the counseling groups met once each week for sessions of about 1¼ hours. These sessions were tape recorded and were observed by research assistants who recorded behavior which might not appear obvious from the tapes. The research assistants also maintained attendance records.

The role of the counselors was never rigidly defined. It was assumed that antecedents of anxiety in college students were likely to be heterogeneous. Furthermore, there appeared to be little agreement in the literature as to the most effective methods for working with anxious students. Therefore, a variety of situation- and person-oriented guidance and counseling procedures was employed. Both counselors attempted to provide a permissive atmosphere in which students' views and feelings could be expressed to stimulate group discussion, to facilitate the utilization by the group of ideas presented by individual students, to furnish relevant factual information when this was requested by members of the group, and to summarize and clarify what was said. On occasion, the counselors pointed out implications of the discussions and suggested applications to new situations.

Students were encouraged to bring up questions or problems of any sort for discussion. Topics of greatest interest included: methods of study; individual academic difficulties; relations with professors in class and on the campus; dormitory life; vocational goals; and, to a lesser extent, matters relating to personal and social adjustment problems. The students wanted to know how to study; how to budget their time; how to prepare for examinations; how to figure out what instructors expected of them; and how to get their work done in dormitories where noise, interruptions, and social demands conspired to distract them. A considerable amount of time was also consumed with students' descriptions of the shortcomings of instructors and courses. At times, feelings of frustration and anger were expressed about the unreasonable expectations of professors. The freshman's low position in the social pecking order was also occasionally discussed.

Although it was expected that a significant portion of the counseling sessions would be devoted to discussions of experiences which had induced anxiety in these students prior to their entering college, the students tended to avoid talking about their personal problems. Instead, as already noted, they were concerned almost exclusively with anxiety-arousing aspects of their present circumstances and with finding ways to cope with the immediate stresses encountered in their new environment. On those infrequent occasions when highly personal matters were brought up, they most often emerged in disguised form. Thus, although the counselors anticipated and were prepared to deal with expressions of more basic personal and emotional difficulties, these did not play a central role in the discussions of the first semester counseling groups. There was, however, abundant evidence of high manifest anxiety in the behavior of the students in the counseling groups, as well as in their descriptions of their reactions to environmental stresses.

Group Counseling Procedures: Second Semester

The main concern of this investigation was in the effect on academic performance of group counseling initiated early in the first semester of the freshman year. The research design, however, included the provision of offering anxious freshmen volunteers assigned to the C group an opportunity to participate in counseling groups during the second semester. Therefore, at the beginning of the second semester students in the C group were invited by letter to participate in seminar groups in which they might discuss academic, vocational, personal, and social problems. In the 1959 sample only 6 students in the C group volunteered; in 1960 10 students in the C group volunteered.

Students in the E group were apprised, at the end of the first semester, that they could continue to participate in the counseling groups during the second semester if they wished to do so. Eight students in the E group in 1959 decided to continue during the second semester, and 3 students did so from the 1960 E group. Thus, a total of 14 anxious freshmen who entered the university in 1959, and 13 who entered in 1960, participated in the second-semester counseling groups. In both years the students were divided into two counseling groups; each counselor worked with one of these groups.

The second-semester counseling groups, in both years, were initiated during the third week of the semester and were conducted along similar lines to those followed during the first semester. They met once each week for about 1¼ hours; each session was tape recorded and observed by a research assistant. Each group met for 10 counseling sessions in 1959 and for either 10 or 11 sessions in 1960. The average number of second-semester counseling sessions attended in both years was 4.1, with a range of 1–8 sessions in 1959 and 2–7 sessions in 1960.

The content of the second-semester counseling sessions, in both years, was quite different from that of the first-semester counseling groups. Although problems and questions concerning study habits, academic difficulties, dormitory life, vocational goals, and the like were raised, the students more frequently, and more readily, discussed problems associated with their personal and social adjustment. This greater freedom in discussing matters of personal concern appeared to result from a number of factors. Included among these were: (*a*) Students who had participated in the counseling sessions during the previous semester felt more comfortable in the counseling groups. (*b*) Many of the situational problems which had initially served to arouse anxiety had already been worked through by the students. (*c*) The counseling groups were smaller, affording those in attendance more opportunity to discuss their personal problems. (*d*) Most of the students from the C group who volunteered appeared to have been motivated to do so because of personal problems which had markedly interfered with their academic adjustment during the previous semester.

Evaluation of the Effects of Group Counseling

Grades were the principal criterion measures of academic adjustment on which anxious freshmen, who participated in counseling groups, were compared with control groups of noncounseled anxious students. Midterm, first-semester, and second-semester grade-point averages (GPAs) were obtained from official university records for each anxious freshman. The student's GPA was the weighted average of his academic performance in course work where 4 points were credited for each hour of A, 3 for B, 2 for C, 1 for D, and 0 for F.

Because of attrition resulting from dropouts (due to transfers, academic failures, and the like), GPA was not considered an appropriate index of the impact of group counseling on academic performance beyond the first year. Therefore, the long-term effects of group counseling were evaluated by comparing the percentages of counseled and noncounseled students who dropped out of school because of academic failure. *Academic failure* was defined for the purpose of this investigation as (*a*) dismissal from the university because of unsatisfactory academic performance, and/or (*b*) leaving the university with a GPA below 1.75 (1.90 was required to graduate). The second, more stringent, criterion was deemed necessary because many students transfer or drop out of school after failing a number of courses; but the students are not officially classified as academic failures if their cumulative GPA at the time that they left was not below the minimum GPA required for them to remain, e.g., only 1.20 and 1.40 for freshmen and sophomores, respectively.

A 2-year follow-up period served to determine the number of counseled and noncounseled anxious freshmen entering the university in 1959 who dropped out because of academic failure before the beginning of their junior year; the follow-up period in 1960 was also 2 years. Differences between counseled and noncounseled students on other variables which might account for any obtained differences in academic performance were also assessed. In addition to the matching variables, data were obtained for the E and C groups on class attendance, the Brown-Holtzman (1955) Survey of Study Habits and Attitudes (SSHA),[3] and the individual MMPI validity and clinical scales. Fraternity rushing and pledging, which regularly take place early in the second semester, were also taken into account in evaluating the effects of second-semester group counseling.

Although the present investigation was concerned primarily with the effects of group counseling on academic performance, supplementary information was obtained through interviews with students in the E and C groups on adjustment in nonacademic areas and on reactions to the project itself. These interviews were conducted during the second semester of the freshman year by project assistants who had not had any previous contact with the students whom they interviewed. Students were asked about their social and extracurricular activities; their educational and vocational plans; and their reactions to the university, the MMPI, and the counseling groups. In 1959 23 of the students in the E group (89%) and 24 of the students in the C group (89%) were interviewed. In 1960 the numbers of students in the E and C groups who were interviewed were 29 (85%) and 28 (74%), respectively. Due to the heterogeneity of the students' responses, and the small cell frequencies which resulted when attendance at group-counseling sessions was taken into account, systematic quantitative analyses of the interviews were not attempted. However, interview data which help to clarify observed relationships between group counseling and academic performance are presented in the context of the discussion of these relationships.

RESULTS

The results of this investigation will be presented in three parts: (*a*) an analysis of the effects of first-semester group counseling on first-semester grades, (*b*) an evaluation of the effects of first-semester counseling on dropouts due to academic failure during the 2-year follow-up period, and (*c*) an analysis of the influence of second-semester group counseling on improvement in GPA.

Effects of First-Semester Group Counseling on Grades

Midterm and first-semester GPAs served as the criteria against which the immediate effects of group counseling were evaluated. Attendance at group-counseling sessions was taken into account, since counseling was expected to have greater influence on the academic performance of students who more regularly attended counseling sessions. The median number of counseling sessions attended in 1959 was 6.5. Students in the 1959 sample, who attended 7 or more first-semester counseling sessions, were designated High Attenders ($N = 13$); those who attended fewer than 7 sessions were designated Low Attenders ($N = 13$). To facilitate comparisons between the 1959 and 1960 samples, High and Low Attenders in 1960 were defined according to the same criterion; 20 students attended 7 or more sessions in 1960 while 14 attended fewer than 7 sessions.

Effects of Group Counseling on Grades for the 1959 Sample. The means and standard deviations of the midterm and first-semester GPAs for the High Attenders, Low Attenders, and Control groups in the 1959 sample are presented in Table 2. For all three groups first-semester GPAs were higher than midterm GPAs. This general improvement from midterm to the end of the semester appeared to reflect the grading practices of many instructors at the university who regularly assigned lower grades at midterm than at the end of the first semester.

Since counseling was not initiated in 1959 until shortly before midsemester, midterm grades provided a direct comparison of the academic potential of the three groups. As may be noted in Table 2, the grades obtained by High Attenders, Low Attenders, and Controls were quite com-

---

[3] The SSHA was given at the beginning of the third group-counseling session in each year and was interpreted and discussed in the following sessions.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF MIDTERM AND FIRST-SEMESTER GRADES FOR HIGH ATTENDERS, LOW ATTENDERS, AND NONCOUNSELED CONTROLS IN 1959

Group	*N*	Midterm		First semester		ACE	
		M	*SD*	*M*	*SD*	*M*	*SD*
High Attenders	13	2.13	.43	2.62	.49	131.5	15.5
Low Attenders	13	2.11	.46	2.23	.50	129.8	15.9
Controls	27	2.20	.59	2.35	.63	132.1	15.1

parable at midterm; and the three groups were also well matched in scholastic aptitude as measured by the ACE. Therefore, improvement in GPA from midterm to the end of the first semester directly reflected the effects of counseling on academic performance. The relative improvement in GPA for High Attenders, Low Attenders, and Controls from midterm to the end of the semester was evaluated by an analysis of variance (Type I, Lindquist, 1953); the results of this analysis are summarized in Table 3. The significant Groups × Grading Periods interaction, when considered along with the means in Table 2, indicated that the High Attenders showed more improvement than did either the Low Attenders or the Controls. The highly significant effect of Grading Period reflects the generality of the grading practice previously noted.

In order to assess more directly the effects of attendance at group-counseling sessions on academic performance, the data for the control group were excluded and the midterm and first-semester GPAs for the High and Low Attenders were evaluated. A significant Groups × Grading Period interaction ($F = 12.16$, $df = 1/24$, $p < .005$) was obtained in this analysis, indicating that the High Attenders improved more than the Low Attenders. A Pearson product-moment correlation was then computed to determine the magnitude of the linear relationship between attendance at counseling sessions and improvement in GPA. The r of .63 ($p < .01$), between number of counseling sessions attended and GPA improvement (first-semester GPA minus midterm GPA), indicated a moderately strong relationship between these variables.

TABLE 3

ANALYSIS OF VARIANCE OF MIDTERM AND FIRST-SEMESTER GRADES FOR HIGH ATTENDERS, LOW ATTENDERS, AND NONCOUNSELED CONTROLS IN 1959

Source	*df*	*MS*	*F*
Groups (G)	2	.261	.452
Error-b	50	.577	
Grading Period (GP)	1	1.303	29.614**
G × GP	2	.300	6.818*
Error-w	50	.044	

* $p < .01$.
** $p < .001$.

Although greater GPA improvement was found for High Attenders than for either Low Attenders or Controls, factors other than counseling might have accounted for both the improvement in grades and the regular attendance at counseling sessions of the High Attenders. Therefore, the High and Low Attenders were compared on all available measures found previously to be related to academic performance. These included the original matching variables, and, in addition, scores on the SSHA and the two anxiety measures. The two groups were also compared on class attendance. The means and standard deviations for High and Low Attenders on measures conceivably related to academic achievement are presented in Table 4. When statistically evaluated no significant differences were found between the High and Low Attenders on any of these measures. Furthermore, comparisons of High and Low Attenders on individual MMPI scales and on MMPI-derived personality profiles also failed to reveal any significant differences between them; these findings have been reported elsewhere (Spielberger, Weitz, & Denny, 1962).

TABLE 4

Means and Standard Deviations for the 1959 High and Low Attenders on Variables Possibly Related to Academic Performance

Variable	Low Attenders[a]		High Attenders[b]	
	M	*SD*	*M*	*SD*
ACE	129.77	15.95	131.54	15.51
% private school	15.38	10.00	15.38	10.00
% declared major	23.08	11.69	15.38	10.00
Class absences	8.85	8.45	5.85	4.24
SSHA[c]	32.00	5.48	34.50	10.81
MA scale	22.00	5.55	20.31	3.34
A scale	18.69	5.58	17.38	5.66

[a] $N = 13$

[b] $N = 13$.

[c] SSHA scores unavailable for one High and five Low Attenders.

Effects of Group Counseling on Grades for the 1960 Sample. The means and standard deviations in 1960 for the midterm GPAs; first-semester GPAs; and CEEB scores of High Attenders, Low Attenders, and Controls are presented in Table 5. It may be noted that the mean GPA at midterm of the High Attenders was higher than that of either the Low Attenders or the Controls. The differences among the three groups in midterm GPAs approached significance when evaluated by a simple randomized analysis of variance ($F = 2.62$, $df = 2/71, p < .10$). Since group counseling was initiated at the beginning of the first semester for the 1960 sample, it was conceivable that counseling had beneficially affected midterm GPAs.

The High Attenders, as may be noted in Table 5, maintained their superiority, but showed less improvement from midterm to the end of the semester than did Low Attenders and Controls. However, since the rank order of the mean GPAs for the three groups was the same as that of their mean CEEB scores, the observed differences in first-semester GPA could have resulted from differences among the groups in scholastic aptitude. Therefore, in order to partial out the effects of group differences in CEEB, the equation for the regression of the two components of the CEEB score on GPA was determined, and mean deviation scores (obtained GPA minus predicted GPA) were computed for High Attenders, Low Attenders, and Controls. The mean deviation scores for the three groups were .16, −.03, and .04, respectively. Although the mean deviation scores for the three groups could be interpreted as indicating that attendance at group-counseling sessions facilitated the academic performance of the High Attenders, an analysis of variance of the deviation scores failed to yield any significant differences among groups.[4] Moreover, the mean deviation score of .04 obtained by the Controls clearly indicated that the noncounseled anxious freshmen were not underachievers.

Effects of First-Semester Counseling on Academic Failures

In evaluating the long-term effects of group counseling on the prevention of academic failure, the data for the 1959 and 1960 samples were pooled. The number of High Attenders, Low Attenders, and Controls who dropped out of school because of academic failure within 2 years of the time these students first entered the university is indicated in Table 6. Although a smaller percentage of High Attenders were academic failures, the difference in the proportion of dropouts due to academic

---

[4] The GPAs of the High Attenders, Low Attenders, and Controls were somewhat more variable in 1960 than during the previous year. This may be noted by comparing the standard deviations for the three groups in 1960 (see Table 5) with those of the comparable groups in 1959 (see Table 2). The greater variability in GPA observed during 1960 may reflect the heterogeneity previously noted in the motivations which prompted students to volunteer during the second year of the project.

TABLE 5

MEANS AND STANDARD DEVIATIONS OF MIDTERM AND FIRST-SEMESTER GRADES FOR HIGH ATTENDERS, LOW ATTENDERS, AND NONCOUNSELED CONTROLS IN 1960

Group	N	Midterm		First semester		CEEB	
		M	SD	M	SD	M	SD
High Attenders	20	2.45	.64	2.48	.71	1235.6	102.8
Low Attenders	14	1.99	.53	2.12	.65	1164.4	111.6
Controls	38	2.14	.64	2.34	.72	1214.3	112.4

failure for the three groups was not statistically significant ($\chi^2 = 2.59$, $df = 2$), indicating that the influence of group counseling during the first semester on preventing academic failures was not clearly demonstrable within a 2-year follow-up period.

Effects of Second-Semester Group Counseling on Grades

The influence of group counseling during the second semester on academic performance was evaluated by comparing second-semester GPAs with those earned in the first semester. If second-semester counseling had beneficial effects on academic performance, counseled students should show greater improvement in GPA from the first semester to the second semester than noncounseled students. The means and standard deviations of first- and second-semester GPAs for anxious freshmen counseled during the second semester, including those who were also counseled during the first semester, and for anxious freshmen who volunteered for the AOP, but who were not counseled in either the first or second semester, are presented in Table 7. It may be noted that students counseled during the second semester of 1959 showed a decrement in GPA in contrast to those counseled in 1960 who improved. An analysis of variance (Type I) of the data from the 1959 sample yielded no significant findings; a similar analysis carried out with the data from the 1960 sample also failed to yield any significant effects.

TABLE 6

ANXIOUS HIGH ATTENDERS, LOW ATTENDERS, AND NONCOUNSELED CONTROLS WHO DROPPED OUT OF COLLEGE BECAUSE OF ACADEMIC FAILURE WITHIN THE FIRST 2 YEARS

Group	N	Academic failures	
		N	%
High Attenders	33	1	3.03
Low Attenders	27	4	14.81
Noncounseled	65	7	10.77

Although it would appear that second-semester group counseling did not produce any systematic improvement in GPA, it was possible that participation in fraternity rushing and pledging influenced grades and that the influence of these uncontrolled variables obscured the relationship between second-semester counseling and academic performance. In both years fraternity rushing began early in the second semester and extended over a 3-week period while the pledge period continued almost until the end of the semester. Participation in these activities required a considerable investment of time and made intensive physical and emotional demands on the student. The mean first- and second-semester GPAs for anxious freshmen who pledged fraternities during the second semester are compared in Table 8 with those of students who did not pledge. In both samples, anxious freshmen who pledged fraternities showed a decline in GPA from the first semester to the second semester; in contrast, those who did not pledge showed corresponding increments in GPA. An analysis of variance of first- and second-semester GPAs for students pledged to fraternities in 1959 and those who did not pledge yielded a highly significant Groups × Grading Periods interaction ($F = 14.28$, $df = 1/50$, $p < .001$). A similar analysis, carried out on the 1960 sample, also yielded a highly significant

TABLE 7

Mean First- and Second-Semester GPAs for Anxious Freshmen Counseled during the Second Semester and for Anxious Freshmen Volunteers Who Were Not Counseled in Either the First or the Second Semester

Group	*N*	First semester GPA		Second semester GPA	
		M	*SD*	*M*	*SD*
1959					
Counseled	14	2.47	.74	2.28	.78
Noncounseled	38	2.36	.51	2.47	.67
1960					
Counseled	13	2.28	.66	2.54	.50
Noncounseled	57	2.35	.73	2.23	.71

Groups × Grading Periods interaction ($F = 18.61$, $df = 1/68$, $p < .001$). Thus, it would appear that rushing and pledging had such pronounced effects on second-semester grades that any beneficial effects which may have been produced by group counseling were obscured.[5]

In order to evaluate the effects of group counseling independent of the influence of fraternity affiliation, the data for the 1959 and 1960 samples were pooled, and the effects of participation in second-semester counseling groups were evaluated for only those students who did not pledge fraternities. The mean first- and second-semester GPAs for the counseled and noncounseled students who did not pledge fraternities are presented in Table 9. It may be noted that these students were closely equated for first-semester GPA, and that counseled students showed more improvement during the second semester than noncounseled students. When these data were evaluated by an analysis of variance, however, the only statistically significant effect was that of Grading Period. The failure to find a significant interaction between Groups and Grading Periods would seem to indicate that participation in the counseling groups during the second semester had little beneficial effect upon academic performance.

Discussion

In this investigation we have attempted to demonstrate that the academic performance of college freshmen, who on the basis of elevated anxiety scores were expected to become academic underachievers or failures, could be facilitated through a special orientation program. This program consisted of weekly group-counseling sessions initiated as early as possible in the students' first semester of college. The design of the investigation also provided group counseling during the second semester for students who participated in the first semester and wished to continue, and for students initially assigned to noncounseled control groups.

First-Semester Group Counseling and Academic Performance

For the anxious freshmen who entered the university in 1959, it was found that those who regularly participated in group-counseling sessions showed greater improvement in GPA from midsemester to the end of the first semester than anxious volunteers who were not counseled or who

---

[5] The average GPAs of students who pledged fraternities were systematically higher than those of nonpledges because the university required a GPA of 2.00 in order for a student to be eligible to pledge. It should be noted, therefore, that the effects of rushing and pledging are to some extent confounded with, and augmented by, regression phenomena, especially in the 1960 sample. When nonpledges with first-semester GPAs below 2.00 were eliminated, the first-semester GPAs of the remaining nonpledges were comparable to those of the fraternity pledges. The significant Groups × Grading Periods interactions ($p < .01$), obtained for both samples when only the grades of those students eligible to pledge were considered, indicated that the decline in the grades of anxious freshmen who pledged fraternities and the improvement of nonpledges were not artifacts of regression phenomena.

TABLE 8

MEAN FIRST- AND SECOND-SEMESTER GPAs FOR ANXIOUS FRESHMEN PLEDGED TO FRATERNITIES DURING THE SECOND SEMESTER AND FOR ANXIOUS FRESHMEN WHO DID NOT PLEDGE A FRATERNITY

Group	N	First semester		Second semester	
		M	SD	M	SD
1959					
Pledged	21	2.55	.36	2.26	.58
Not pledged	31	2.28	.67	2.53	.75
1960					
Pledged	32	2.70	.44	2.36	.67
Not pledged	38	2.03	.76	2.33	.69

irregularly attended counseling sessions. Furthermore, improvement in GPA was significantly correlated with the number of counseling sessions attended. There was no evidence in available biographical and personality data of any common factor which might have accounted for both the improvement in grades and the regular attendance at counseling sessions for those students who benefited from participation in the counseling group.

Although it would seem reasonable to conclude that regular attendance at group-counseling sessions facilitated the academic performance of the 1959 High Attenders, just how these students had been helped by their counseling experiences was not at all clear from what went on in the counseling sessions. The interviews, conducted during the second semester with anxious freshmen who had participated in the first-semester counseling groups, shed some light on this subject. In the interviews the students were asked to indicate specific ways in which they felt that the group sessions had been helpful. They were also asked to indicate any complaints or disappointments related to their experiences in the counseling groups. The particular way in which the group sessions had been helpful, which was most frequently mentioned by both High and Low Attenders, was that they provided the student with an opportunity to see that other students had problems similar to their own. The next most frequently cited, helpful aspect of the counseling sessions, for both High and Low Attenders, was the benefit derived from hearing about other students' experiences and opinions. It would seem *unlikely*, however, that these factors differentially accounted for the greater improvement in grades of the High Attenders, inasmuch as they were mentioned equally often by the Low Attenders.

In contrast to Low Attenders, a majority of the High Attenders indicated that they felt that the group sessions had helped them in their approach to studying and exam taking, and in their relationships with their professors. Only a few Low Attenders stated that the counseling sessions helped them in these specific ways. High Attenders also tended to emphasize ways in which the group-counseling experience helped them, whereas Low Attenders more often voiced complaints. On the basis of their interview responses and

TABLE 9

MEAN FIRST- AND SECOND-SEMESTER GPAs FOR COUNSELED AND NONCOUNSELED ANXIOUS FRESHMEN WHO DID NOT PLEDGE FRATERNITIES

Group	N	First semester		Second semester	
		M	SD	M	SD
Counseled	19	2.14	.69	2.52	.69
Noncounseled	50	2.15	.75	2.38	.74

the quality of their participation in the group sessions, the High Attenders appeared to be more "psychologically-minded" and more interested in other students than were the Low Attenders who seemed primarily to be concerned with deriving specific assistance in getting better grades.

The attempt to replicate, in 1960, the finding that group counseling had beneficially influenced the academic performance of anxious freshmen who entered the university in 1959, was complicated by apparent alterations in the university environment. Most of the students who volunteered for the AOP in 1960 had read favorable reports about the special orientation program in the campus newspaper. In addition, some students had heard about the counseling groups from friends who had previously participated in them. These factors led to increased student interest in the AOP in 1960 which was reflected to some extent in the percentage of students invited who volunteered to participate, and, more strikingly, in the promptness of their volunteering. Of the anxious freshmen who volunteered in 1959, only 44% responded to the initial invitation; in 1960, the percentage of the anxious volunteers responding to the initial letter was 88%.

There were differences in the general level of academic aptitude of the students who entered the university in 1959 and in 1960, and these differences were reflected in the mean CEEB scores of anxious freshmen. The CEEB scores (V plus M) for the anxious freshmen in 1959 and in 1960 were 1,149.6 and 1,190.5, respectively. The relationship between anxiety and academic performance as a function of academic aptitude also appeared to be different for the two samples and altered from what it had been in previous years. Differences in the two samples in the level of academic aptitude and in the relationship been anxiety and grades appeared to be related to discernible trends in university student-selection procedures and will be discussed below.

The findings for the 1960 sample, with respect to the effects of first-semester group counseling on the grades of anxious freshmen, are difficult to interpret. Although the grades of anxious students who regularly attended counseling sessions during the first semester of 1960 were higher, both at midterm and at the end of the semester, than were those of either Low Attenders or noncounseled Controls, the differences among the three groups, when statistically evaluated, were not found to be significant. Moreover, the observed differences in favor of High Attenders could have resulted in part from the fact that the High Attenders had higher CEEB scores than Low Attenders and Controls. Thus, although first-semester group counseling in 1960 appeared to have some beneficial influence on the grades of anxious freshmen who regularly participated in the counseling groups, the counseling effect was not strong enough for statistical tests to reach a satisfactory level of significance and were confounded with differences among the groups in scholastic aptitude.

Of the total of 33 High Attenders in the first-semester counseling groups in 1959 and 1960, only one student (3%) dropped out within a 2-year follow-up period because of academic failure. The percentages of dropouts for Low Attenders and for noncounseled Controls were approximately 15% and 11%, respectively. Although the percentage of academic failures was less for the anxious freshmen who regularly participated in the counseling groups, differences in the percentage of dropouts among the three groups were not statistically significant. It was apparent that a longer follow-up period would be required in order to determine whether or not first-semester group counseling helped to reduce significantly the number of anxious freshmen who eventually dropped out of college because of academic failure.

Second-Semester Group Counseling and Academic Performance

Considering the number of anxious freshmen who volunteered for the AOP at the beginning of the first semester, there was surprisingly little interest in the second-semester counseling groups. Of the students who volunteered in 1959 and in 1960, and who were subsequently assigned to the Control group, only 16 (25%) were interested in participating in counseling groups during the second semester. Several of these students had done very poorly during their initial semester and were in imminent danger of being dropped from school for academic failure.

A similar lack of interest in the second-semester counseling groups was evidenced by students who had participated during the first semester. Only 11 students (18%) who had taken part in first-semester counseling groups desired to continue in the second semester. Nine of these were High Attenders in the first semester, who had performed at a relatively high academic level. For these students, the possibility of further improvement was limited by ceiling effects and regression to the mean phenomena.

The heterogeneity apparent in students' motivations for participating in second-semester counseling groups was reflected in attendance at counseling sessions. During both 1959 and 1960 the mean number of sessions attended was only 4.1 with a range from 1 to 11 sessions for individual students. The majority of the students who participated in second-semester counseling groups attended no more than 3 sessions. However, four students attended 10 or more sessions. In sum, only a small number of students participated in the second-semester counseling groups; and their apparent motivations for participating, their academic performance in the previous semester, and their attendance at counseling sessions were quite varied. The content of the second-semester counseling sessions, however, was more like that usually encountered in group counseling and group therapy with college students than was the case in the first-semester groups. In the second-semester groups the students more readily discussed problems associated with their personal and social adjustment, whereas academic problems characterized the discussions in the first-semester counseling groups.

There was no clear evidence that the second-semester counseling groups beneficially influenced the academic performance of anxious students. There was convincing evidence, however, of the adverse effects of fraternity rushing and pledging on second-semester grades. Irrespective of participation in group counseling, anxious freshmen who pledged fraternities showed a marked decline in GPA from the first to the second semester, while those who did not join fraternities made better grades in the second semester than they had achieved during the first semester. Although the discovery that an environmental factor such as fraternity affiliation could have a marked impact on the behavior of individuals receiving counseling or psychotherapy was perhaps not surprising, it was unusual for this effect to be so clearly and consistently demonstrated. The influence of uncontrolled environmental factors on counseling results has been frequently acknowledged, but seldom examined, except in individual case studies.

The adverse effect of fraternity affiliation on academic performance was strikingly demonstrated for the four anxious freshmen who were High Attenders in the first semester of 1959, and who pledged fraternities while continuing to participate in counseling groups during the second semester. The mean first-semester GPA of these students was 3.12 as compared to a mean GPA of only 1.91 for the second semester. Decrements in GPA for individual students ranged from −.53 to −1.88. It should be noted, moreover, that the mean GPA of 1.75 earned by these four students during the first semester of their sophomore year was even lower than the grades they earned during the semester in which they participated in fraternity rushing and pledging. Thus, the stresses and time commitments associated with fraternity membership not only outweighed any possible beneficial influences of second-semester group counseling, but also tended to undo the positive effects of participation in the first-semester counseling groups.

As a group, the anxious freshmen who did not pledge fraternities made higher grades in the second semester of their freshman year than they had achieved during the first semester, irrespective of their participation in the counseling groups. Of the anxious freshmen who did not pledge fraternities, however, those who participated in the second-semester counseling groups showed more improvement in GPA from the first semester to the second semester than did noncounseled, nonfraternity students. The magnitude of the difference in improvement for counseled and noncounseled students was not statistically significant, however, suggesting that even when fraternity membership was taken into account, second-semester counseling still had little beneficial effect on academic performance. The small percentage of students who volunteered for the second-semester counseling groups, the low attendance at group sessions of those who did participate, and the failure to demonstrate any relationship between attendance at counseling sessions and academic performance, would seem to indicate that group counseling with anxious freshmen during their second semester at college does not provide a meaningful approach to facilitating the academic performance of this kind of potential underachiever.

Relationship between Anxiety and Academic Performance

The assumption in the present investigation that anxious college freshmen would become underachievers if left to their own resources was based upon previous findings in which it was demonstrated that anxiety had adverse effects upon the academic performance of college upperclassmen (Katzenmeyer, 1958; Spielberger, 1962; Spielberger & Katzenmeyer, 1959). In these studies: (*a*) Anxious students in the middle ranges of academic ability obtained significantly lower grades and had a higher percentage of academic failures than nonanxious students of comparable ability. (*b*) Students at the very lowest ability level earned low grades irrespective of their anxiety level, but a higher percentage of anxious, low-ability students eventually dropped out of school because of academic failure. (*c*) For a small number of students of very superior ability, it appeared that anxiety actually facilitated academic performance. In the present investigation, the relationship between anxiety and grades for male liberal arts freshmen appeared to be altered from what it had been in the original studies. It also appeared to be somewhat different for students entering the university in 1959 and those entering in 1960.

The relationship between anxiety and first-semester GPA for anxious freshmen who entered the university in 1959 and 1960, excluding those who participated in first-semester counseling groups, are compared in Figure 2 with the original findings for upperclassmen (Spielberger, 1962). The upperclassmen had been given the *MA* scale at the beginning of the semester in which they were enrolled in the introductory psychology course, and their GPAs were determined for that same semester. Several important differences may be noted in the relationships depicted in Figure 2 between anxiety and grades as a function of academic aptitude.[6] First, for students of low aptitude, high anxiety appeared to: (*a*) impair the academic performance of the 1954–57 upperclassmen, (*b*) have little effect on the performance of the 1959 freshmen, and (*c*) facilitate the performance of the 1960 freshmen. Second, for average aptitude students, although high anxiety led to lower performance in each of the three samples, the extent to which anxiety impaired performance was con-

---

[6] The criteria for the high- and low-anxiety groups and levels of scholastic aptitude varied slightly from sample to sample. For all three samples, however, the high- and low-anxiety groups represented approximately the upper and lower 20% of the distributions of scores on the Taylor *MA* scale for the particular sample. Low-anxiety upperclassmen had *MA*-scale scores of 7 or less; low-anxiety freshmen had scores of 5 or less in 1959 and of 6 or less in 1960. For all three samples, high-anxiety students had *MA*-scale scores of 19 or above. Scholastic aptitude levels for the upperclassmen and the 1959 freshmen were based on local stanine norms (ACE total scores) for entering freshmen for the period, 1951–55. The aptitude levels in 1960 were based upon CEEB stanines for freshmen entering the university in the 3-year period, 1957–59. The ACE or CEEB stanines represented by the Lo., Avg., and Hi. aptitude levels in Figure 2 were 1–4, 5–6, and 7–9, respectively.

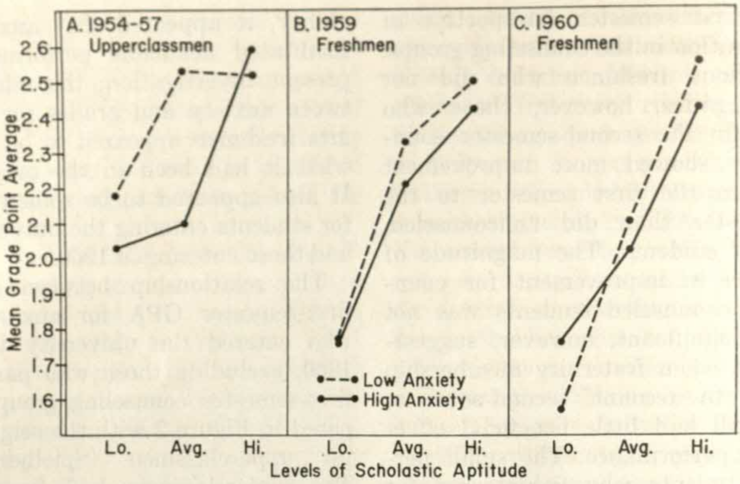


FIG. 2. The relationship between anxiety and grades as a function of scholastic aptitude for: (A) upperclassmen (1954–57), (B) freshmen who entered the university in the fall term of 1959, and (C) freshmen who entered in the fall term of 1960.

siderably less for both groups of freshmen than was the case for the upperclassmen. Third, for the students of high aptitude, high anxiety tended to facilitate the academic performance of the upperclassmen, but led to impaired performance in both freshmen samples. Before attempting to account for these differences in the relationship between anxiety and grades, changes in the intellectual characteristics of students admitted to the university over the past decade need to be noted.

Between 1953 and 1960 the average level of academic aptitude of male freshmen entering the university has been consistently higher each successive year. The percentile equivalent (national norms) of the mean ACE total score for male liberal arts freshmen who entered the university in the fall of 1954 was the sixty-seventh percentile. In 1959, the last year in which the ACE test was given, the percentile equivalent for the mean ACE score for entering male liberal arts freshmen was the seventy-ninth percentile. This improvement in the average aptitude level of students entering the university has resulted from the systematic exclusion of applicants of lower ability (Spielberger, 1958). But how does the higher ability level of students entering the university influence the relationship between anxiety and grades? The answer to this question is complex and requires prior consideration of the effects of anxiety on tasks which vary in difficulty.

It has been demonstrated that high anxiety (*a*) facilitates performance on easy tasks, i.e., tasks in which a correct response tendency is clearly dominant; and (*b*) impairs performance on more difficult or complex tasks, i.e., those in which strong, incorrect responses are elicited which compete with the correct response tendency (Spence, 1958; Spielberger, Goodstein, & Dahlstrom, 1958). Assuming that the academic tasks from which grades are determined are relatively complex, and that the difficulty level of the academic task varies inversely with a student's ability, it might be expected on the basis of previous findings that anxiety would facilitate the academic performance of average and low ability students.

The tendency for anxiety to facilitate the academic performance of upperclassmen with high ability may be noted in Figure 2. The facilitative effects of anxiety were even more pronounced when only the grades of the most able students were considered.[7] By 1959, however, it would appear that the academic task at the university was no longer easy even for high ability students. It may be seen in Figure 2 that high anxiety led to impaired academic performance for high ability freshmen in both 1959 and 1960. This may be explained if it is noted that although the average ability level of students at the university has improved markedly, the average grades awarded to students tend to be about the same as in the past. Thus, higher standards for admission and the matriculation of more able students have contributed to increasing the difficulty level of the academic task and reversing the effects of anxiety on grades for high ability students.

---

[7] For upperclassmen with ACE scores of 150 or higher, the mean GPA for high-anxious students was 3.01 as compared to a mean GPA of 2.70 for low-anxious students of comparable ability (Spielberger, 1962).

An even more complex explanation would seem to be required in order to account for the academic performance of freshmen of relatively low ability, especially for the surprising finding in 1960 that high-anxious students of low ability did better than low-anxious students of comparable ability. To account for the latter finding, it is essential to take cognizance of factors which figure importantly in the admission of students to the university. Applicants for admission are judged on the basis of (*a*) their scores on objective tests of scholastic aptitude, principally the CEEB; (*b*) their academic performance in high school, taking into account the quality of the high school attended; and (*c*) recommendations of their high-school teachers and representatives of the university by whom they are interviewed. To gain admission to the university, students with relatively low scores on the CEEB must have good high-school grades and strong recommendations.

On the basis of previous findings, however, it would be expected that high anxiety would have adverse effects on the high-school grades of candidates for admission to the university with relatively low aptitude (low CEEB scores). For an anxious student with low aptitude (but perhaps average CEEB scores for high-school students), good high-school grades and recommendations would be expected only if the student had developed superior study habits which allowed him to achieve despite his anxiety and modest ability. The development of superior study habits, however, would not only lead to good high-school grades, but possibly also to better performance in college than might be expected on the basis of the student's CEEB score. The observation in 1960 that high anxiety facilitated the academic performance of college freshmen of relatively low aptitude (see Figure 2C) would be explained if, as the university's academic standards increased, only high-anxious, low aptitude students with superior study habits were admitted. In an unpublished study of students enrolled in introductory psychology at the university, it was found that high-anxious students of low aptitude had significantly higher scores, indicating better study habits, on the Brown-Holtzman SSHA, than low-anxious students of comparable aptitude. Most of the subjects for whom this relationship was found had entered the university as freshmen in the fall of 1960.

Relationship between Volunteering and Academic Performance for Anxious College Freshmen

The academic performance of high- and low-anxious freshmen with average and superior ability in both 1959 and 1960 was consistent with the assumption that many college freshmen suffer academic underachievement and failure because of anxiety. The performance of freshmen of low aptitude in both samples, however, was not consistent with this assumption as may be noted in Figures 2B and 2C. Of the high-anxiety freshmen who entered the university in 1959, and on whom the data of Figure 2B were based, approximately one third had volunteered for the AOP and were subsequently assigned to the noncounseled Control group during the first semester. The remaining anxious freshmen, whose grades are reflected in Figure 2B, had not volunteered. The 1960 anxious freshmen, whose first-semester grades are depicted in Figure 2C, consisted of approximately equal numbers of noncounseled volunteers and nonvolunteers. The question arises as to the relative contributions of volunteers and nonvolunteers to the observed relationship between anxiety and academic performance.

The mean first-semester GPAs for noncounseled volunteers and nonvolunteers in 1959 were 2.35 and 2.15, respectively. In 1960 the mean first-semester GPAs for the comparable groups were 2.34 and 2.18, respectively. Although the volunteers in both years achieved higher first-semester grades, they also had slightly higher scores on the CEEB. Therefore, differences in academic aptitude may have determined the superior performance of the volunteers. When the relationship between volunteering and GPAs was examined, taking into account the students' academic aptitude by calculating deviations from GPAs predicted from academic aptitude measures, there remained a consistent, albeit nonsignificant, tendency for the anxious freshmen who volunteered for the special orientation program to earn higher grades.

There were also proportionately fewer severe underachievers among the anxious volunteers than among the anxious nonvolunteers. Defining *severe underachievement* as a deviation of −1.00 or greater in obtained GPA from that predicted from aptitude scores, only 3.7% of the noncounseled anxious volunteers were classified as severe underachievers in 1959 as compared to 11.3% of the nonvolunteers. In 1960 5.3% of the noncounseled volunteers were severe underachievers as compared to 10.6% of the nonvolunteers. Of the anxious noncounseled volunteers who entered the university in 1959, 11.1% dropped out of school by the end of their sophomore year because of academic failure, as compared to 21.8% of the nonvolunteers. In 1960 8.0% of the volunteers dropped out of school because of academic failure as compared to 10.6% of the nonvolunteers. Thus, there was a consistent tendency for the noncounseled anxious volunteers to do better in their academic work than anxious nonvolunteers.

Contrary to expectation, the failure dropout rate for anxious volunteers in 1959 was actually less than that for nonanxious students, and in 1960 the failure dropout rates for both anxious volunteers and anxious nonvolunteers were lower than those for nonanxious students. The latter finding was consistent with the data reported in Figure 2C in which it may be noted that high-anxious students with low ability did better than low-anxious students with comparable ability. Since a larger proportion of the students who dropped out because of academic failure during the first 2 years of college came from those with low aptitude, it would appear that admission to the university of high-anxious, low aptitude students with good study habits contributed to the reversal of the previously observed relationship between anxiety and academic failure (Spielberger, 1962).

In sum, the relationship between anxiety and academic performance would appear to be somewhat different for students willing to volunteer for a special academic orientation program than for anxious students who do not volunteer for such programs. Anxious freshmen nonvolunteers tend to earn lower grades, to have a higher proportion of serious underachievers, and to manifest a higher dropout rate due to academic failure, than noncounseled anxious volunteers. Moreover, the noncounseled anxious freshmen who volunteered for the AOP were not underachievers; they did about as well in their first semester as nonanxious students.

Group Counseling as an Approach to the Prevention of Underachievement

In the present study anxious freshmen who volunteered for first-semester counseling groups and regularly participated in them tended to make higher grades than anxious freshmen who volunteered, but who either failed to participate on a regular basis or who were assigned to noncounseled control groups. However, only 50–60% of the anxious students invited to take part in the special orientation program volunteered; and only one third to one half of the volunteers attended a sufficient number of counseling sessions to benefit from the program. Furthermore, contrary to expectation, noncounseled anxious volunteers were not underachievers; they did about as well as nonanxious students. These findings would seem to indicate that a voluntary special orientation program employing group counseling procedures does not provide an effective approach to the prevention of academic underachievement because students who subsequently become underachievers do not volunteer for such programs.

The observation in the present study that anxious freshmen who did *not* volunteer for the special orientation program tended to become academic underachievers was consistent with previous findings that underachievers were reluctant to seek or accept help (Blaine & McArthur, 1961; Walters, 1961). Given the reluctance of the potential underachiever to seek help, the task of assisting such students to achieve up to their capacity appears formidable. If underachievers are to be aided in making more effective adjustments to college life, it is first necessary to develop effective procedures for establishing contact with them.

One potentially promising approach to the problem of underachievement is to work with representatives of the university who normally come into "official" and relatively frequent contact with freshmen at the beginning of their college careers. Psychologists could contribute: (*a*) by helping faculty advisors and dormitory counselors, who generally serve as the principal agents of the university for assisting freshmen to adjust to campus life, to identify students who are likely to need special help; and (*b*) by providing consultation and assistance to these university representatives in their work with individual students. To the extent that university personnel with assigned responsibilities for working with freshmen can be aided in recognizing underachievers and working more effectively with them, many students whose problems might otherwise be ignored may be helped to make better academic, social, and personal adjustments to college life.

SUMMARY

College freshmen who, on the basis of high *MA* scores were expected to become underachievers, were invited to participate in a special academic orientation program. Students who volunteered for this program were assigned to experimental and control groups matched with respect to scholastic aptitude and other variables related to academic performance. The experimental groups were seen weekly in group-counseling sessions throughout the first semester of their freshman year and were invited to continue to participate in the counseling groups during the following semester. Members of the control group were not counseled during the first semester, but were offered the opportunity to participate in counseling groups during the following semester.

Those anxious students who regularly attended group-counseling sessions during the first semester of the freshman year tended to make higher grades than students who either did not regularly attend counseling sessions or who were assigned to control groups during the first semester. Interviews with students who participated in the first-semester counseling groups provided qualitative information on aspects of the counseling process which appeared to influence academic performance. In a 2-year follow-up study it was found that a smaller percentage of the anxious freshmen who regularly attended first-semester counseling groups dropped out of school because of academic failure.

The effects of group counseling during the second semester on the grades of anxious freshmen tended to be obscured by the adverse influence of fraternity rushing and pledging on academic performance. Anxious freshmen who pledged fraternities showed a decline in grades from the first semester to the second semester irrespective of whether or not they participated in counseling groups. Conversely, anxious freshmen who did not pledge fraternities showed corresponding increments in academic performance. Anxious freshmen who participated in counseling groups, and who did not pledge fraternities, tended to show more improvement during the second semester than noncounseled, nonfraternity students.

The effects of systematic trends in university admissions policies on the underlying relationship between anxiety, scholastic aptitude, and academic performance were discussed. Possible differences in the relationship between anxiety and academic performance for students who volunteered for the special orientation program and for nonvolunteers were also explored.

REFERENCES

BAYMUR, F. B., & PATTERSON, C. H. A comparison of three methods of assisting underachieving high school students. *J. counsel. Psychol.*, 1960, **7**, 83–89.

BLAINE, G. B., JR., & MCARTHUR, C. C. *Emotional problems of the student.* New York: Appleton-Century-Crofts, 1961.

BROEDEL, J., OHLSEN, M., PROFF, F., & SOUTHARD, C. The effects of group counseling on gifted underachieving adolescents. *J. counsel. Psychol.*, 1960, **7**, 163–170.

BROWN, W. F., & HOLTZMAN, W. H. A study-attitudes questionnaire for predicting academic success. *J. educ. Psychol.*, 1955, **46**, 75–84.

KATZENMEYER, W. G. A study of the relationship between a personality variable and academic performance. Unpublished master's thesis, Duke University, 1958.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPIELBERGER, C. D. On the relationship between anxiety and intelligence. *J. consult. Psychol.*, 1958, **22**, 220–224.

SPIELBERGER, C. D. The effects of manifest anxiety on the academic achievement of college students. *Ment. Hyg.*, 1962, **46**, 420–426.

SPIELBERGER, C. D., GOODSTEIN, L. D., & DAHLSTROM, W. G. Complex incidental learning as a function of anxiety and task difficulty. *J. exp. Psychol.*, 1958, **56**, 58–61.

SPIELBERGER, C. D., & KATZENMEYER, W. G. Manifest anxiety, intelligence, and college grades. *J. consult. Psychol.*, 1959, **23**, 278.

SPIELBERGER, C. D., WEITZ, H., & DENNY, J. P. Group counseling and the academic performance of anxious college freshmen. *J. counsel. Psychol.*, 1962, **9**, 195–204.

TAYLOR, JANET, A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WALTERS, P. A., JR. Student apathy. In G. B. Blaine, Jr., & C. C. McArthur, *Emotional problems of the student.* New York: Appleton-Century-Crofts, 1961.

WEITZ, H., CLARKE, MARY, & JONES, ORA. The relationship between choice of a major field of study and academic preparation and performance. *Educ. psychol. Measmt.*, 1955, **15**, 28–38.

WEITZ, H., & WILKINSON, H. JEAN. The relationship between certain nonintellective factors and academic success in college. *J. counsel. Psychol.*, 1957, **4**, 54–60.

WELSH, G. S. Factor dimensions A and R. In G. S. Welsh, & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

(Received June 15, 1963)

Vol. 78, No. 14 Whole No. 591, 1964

Psychological Monographs: General and Applied

SEX DIFFERENCES IN BODY PERCEPTION[1]

SEYMOUR FISHER

Upstate Medical Center, State University of New York, Syracuse

A series of studies was undertaken to ascertain whether there are sex differences in body concept and body perception which go beyond the usual phallic-nonphallic distinction. The following parameters were investigated: (a) ability to tolerate aniseikonically induced distortions in the appearance of one's legs, (b) definiteness of one's body image boundaries, (c) the organization of body awareness, (d) accuracy of perception of alterations in the sex of one's appearance, and (e) mode of response to lens-induced alterations in the vertical and horizontal axes of one's mirror image. Multiple samples of college students were used as Ss. The results demonstrated that there are a number of significant sex differences in body image. There was also converging evidence that women may have a more clearly articulated and stable body concept than men.

DESPITE the obvious anatomical distinctions between the sexes, it has not been a simple matter to demonstrate that they differ fundamentally in the organization of the body scheme. That is, aside from the fact that the male and female have unlike body experiences related to their anatomical and body function differences we do not have empirical evidence that their body concepts are distinguished in a more generalized psychological sense. There have been various psychoanalytic speculations concerning this matter. Freud (1938, 1959), for example, suggested that the male with his externally exposed genitals is particularly likely to experience castration anxiety; and he assigned considerable importance to this fact in his explanation of how the Oedipus complex is resolved. Similarly, Freud underscored the fact that the female lacks an externally projecting genital and that the vagina, because it is literally a break in the body wall, lends itself to fantasies that a previously projecting genital has been removed. He theorized that this sense of loss may take the form of penis envy and compensatory attempts to behave in a phallic fashion. Eriksen (1951) reported that when male and female adolescents were asked to construct scenes with play materials, the males tended to fashion upright, phallic structures and the females to depict configurations which are open and easy to enter. He proposed that these differences were actually symbolic expressions of the genital differentiation between the sexes and their contrasting modes of experiencing their bodies. Franck and Rosen (1949) found, in the same vein, that when males and females were asked to complete a series of vague line drawings, the males were typified by completions with projecting phallic qualities. This was interpreted, at least partially, as an expression of sex differences in body feeling.

All of the speculations and reports cited above share the generalization that the male and female are distinguished by body attitudes which are specifically a function of the phallic-nonphallic factor. However, there have also been other lines of inquiry about sex differences in body attitude which should be enumerated. Katcher and Levin (1955) reported that when boys and girls were asked to construct a series of human forms representing self, mother, and father by means of schematic body parts differing in size, the girls seem to have an earlier realistic appreciation than boys of the smallness of their own bodies in relation to those of adults. It was speculated that since body smallness is a characteristic associated with femininity in our culture, this difference might indicate that girls learn more quickly than boys to assign sexual characteristics to their bodies. Somewhat relatedly, Jourard and Secord (1955) observed that adult women differ from men in preferring their body proportions to be small rather than large. Finally several studies should be mentioned which suggest that women may generally have greater awareness of their bodies than do men. Secord (1953) and also Weinberg (1960) reported that when associations to homonyms with body and nonbody meanings were obtained, women gave more body oriented responses than did men. Van Lennep (1957) has similarly noted that in Thematic-Apperception-Test–type stories females exceed males in number of references to body sensations and feelings.

---

[1] This study was partially supported by United States Public Health Service Grant M05761.

One can see that the body scheme differentiation between males and females may not be entirely (or perhaps even primarily) a function of the phallic-nonphallic dimension. Indeed, it is probable that it does involve more generalized variables like feelings about body size and overall awareness of one's body.

The present project was concerned with uncovering the extent of sex differences in body image that do not have direct phallic-nonphallic connotations. It was the objective to explore such differences from several perspectives not primarily derived from the simple fact of unlike anatomy and body function, but rather related to contrasts between the male and the female in their style of life and role in the culture. While the studies to be described are diverse, they are unified by the intent to ascertain promising dimensions of sex differences in body concept. To demonstrate meaningful sex differences in body perception is a necessary preliminary to more systematic ventures in this area. The specific body image dimensions chosen for consideration were those for which there were already initial data suggestive of sex differences.

An important secondary goal of the project was to examine critically assumptions which have been made that women are less satisfied with their bodies than men and also less able to arrive at an articulated, realistic body concept. These assumptions have been fostered not only by stereotypes concerning male superiority, but also the extensive theorizing of Freud and other psychoanalysts who have emphasized that the female is severely handicapped in forming an adequate body concept by the inferiority feelings related to her lack of a penis.

STUDY 1

Aniseikonic Perception of One's Legs

This first study was partially suggested by the work of Wittreich and Grace (1955). They compared male and female subjects of varying ages with respect to the changes they reported in their own appearance while observing their mirror images through aniseikonic lenses.[2] Their data indicated a trend for girls to report, among other things, fewer changes in their legs and feet than was true for boys. It has been shown in several studies that perceptual targets which arouse anxiety or disturbance in an individual are those which he is least likely to see as changed when viewing them through aniseikonic lenses. Perception of change in the appearance of an aniseikonically viewed stimulus seems to be facilitated if that stimulus is experienced as nonthreatening (Wittreich & Radcliffe, 1955). With this perspective, Wittreich and Grace (1955) explained the sex differences they observed in frequency of aniseikonic change in one's own body parts as a function of differential anxiety about such parts. In discussing their data, they particularly focused upon the relative infrequency with which girls perceived changes in their legs as compared to boys and theorized that this was due to the difference in meaning attached to their legs. As they conceptualized it, the boy experiences his legs primarily as a means of locomotion and is frequently able to test out their effectiveness in this respect. Therefore, he has no special anxiety or uncertainty about them. However, they note that for the girl her legs are not only a means of locomotion but also important objects of display that evoke reactions from others which affect her judgment of whether she is attractive. Presumably, then, she would be somewhat dependent upon the opinions of others for arriving at an evaluation of her legs. In that sense, her perception of her legs would be marked by relative uncertainty and anxiety. This explanation was quite speculatively offered by Wittreich and Grace to illustrate the sort of variable that might cause a specific body area to be resistive to perceptual alteration. What is of special pertinence to the present study is their observation that females are more resistive then males to seeing changes in their legs. Subsequent to this work two papers appeared which further suggested that males and females differ in their attitudes toward their legs. Bennett (1960) reported that when subjects were asked to list any 10 body parts, men more often mentioned the leg region (in terms of reference to the knee) than did women. Calden, Lundy, and Schlafer (1959) noted that when male and female college students rated their degree of satisfaction with various parts of their bodies, the women were significantly more critical of their legs and the lower sectors of their bodies than were the men. When the Wittreich and Grace; Bennett; and Calden, Lundy, and Schlafer studies are considered together they imply that men are less dissatisfied with their legs than women and also more freely aware of them and better able to tolerate their perceptual alteration.

---

[2] It should be indicated that aniseikonic lenses result in the image of an object which is formed in one eye differing in size and shape from the image of the same object formed in the other eye.

With this perspective, it seemed promising to theorize that males and females do differ fundamentally in the way they experience their legs. However, the Wittreich and Grace view that this difference in attitude is due to the girl's special dependence upon the reactions of others for arriving at an evaluation of her own legs did not seem to be a sufficient explanation. Thus, one could argue that the boy too is dependent upon the responses of others in arriving at attitudes toward his legs. This dependence would not be so much in terms of reactions to the appearance of his legs as with reference to the degree of agility and skill he shows when performing in games or other athletic-type activities. The boy does not abstractly know that he is skillful in movement. He needs the positive response of others to establish that such is true. What will be suggested here is that the difference between the sexes in attitudes toward one's legs is an expression of contrasting attitudes toward mobility and movement in space. The role of the female is typically defined as less mobile and more "stay put" than is the role of the male. The male is encouraged to learn to move his body aggressively and to venture forth without anxiety. Contrastingly, it is the custom to be more protective of the female as she moves out from the family circle. She is given the feeling that it is less acceptable and more dangerous for a woman to move about freely "out there" than it is for a man. Talcott Parsons (Parsons & Bales, 1955) and others have described in detail the division of labor between the sexes, such that the female's prime duties are within the family group and the male's involve making contact for the family with outside agents and institutions. In short, socialization experiences and expectations make the female feel insecure about the type of mobility and movement which are equated with leg activity; whereas the male is encouraged to regard such mobility as a normal part of his life. It is this fact which will be tentatively singled out as most important in accounting for sex differences in leg attitudes which have been observed.

The primary purpose of the present study was to attempt to replicate with adult subjects the sex differences in aniseikonic perception of the legs which Wittreich and Grace reported for children. Such a replication would solidly establish that a basic sex difference in leg perception exists not only in children but also in adults.

A second prime objective was set which was dependent upon the results obtained with the aniseikonic procedure. It was reasoned that if it could be shown that women do, indeed, feel more insecure than men about using their legs in an active motoric fashion it should be possible to demonstrate that their legs are less physiologically activated than are those of men. This proposition is a logical derivative of previous work in which it has been demonstrated that specific attitudes toward given body areas may be accompanied by particular levels of physiological activation of those areas. Such correlates have been shown for exterior versus interior, right versus left, and front versus back body sectors (Davis, 1960; Fisher, 1958, 1961b, 1963; Fisher & Cleveland, 1958; Shipman, Oken, Grinker, Goldstein, & Heath, in press).

Procedure

Aniseikonic judgments were obtained by asking each subject to stand 48 inches from a full-length mirror and to examine his image while wearing a set of aniseikonic lenses (right eye 4% meridional afocal, X 135°; left eye 4% meridional afocal, X 45°). He was simply told to note any changes in his appearance that might be produced by the lenses. After he had had 5 minutes to observe himself he was told,

> I will read you a list of body areas. As I mention each one, tell me if it was changed in any way by the glasses.

The following list of body areas was then read to the subject: face, shoulders, chest, arms, hips, legs.

Subjects

Two separate samples were evaluated. One consisted of 80 subjects (30 men, 50 women); and the other of 63 subjects (30 men, 33 women). They were all college students recruited by payment of a fee. The median age in both groups was 21.

Results

Sample 1. As shown in Table 1, the male and female subjects in Sample 1 are differentiated in the same fashion as the boys and girls in the Wittreich and Grace sample. There is a significant trend for the men to exceed the women in perceiving aniseikonically induced changes in their legs ($\chi^2 = 4.8$, $p < .02$, one-tailed test). None of the sex differences for the other body parts are significant, although there is a borderline tendency for women to perceive more changes in the face ($\chi^2 = 3.5$, $p < .10$, two-tailed test) and in the chest

TABLE 1

ANISEIKONIC CHANGES IN BODY APPEARANCE REPORTED BY MEN AND WOMEN IN TWO SAMPLES

Body part	Change[a]	Sample 1			Sample 2		
		Male	Female	χ^2	Male	Female	χ^2
Face	Yes	14	34	3.5	14	25	5.4[b]*
	No	16	16		16	8	
Shoulders	Yes	14	18		18	14	1.7
	No	16	32		12	19	
Chest	Yes	12	30	3.1	14	15	
	No	18	20		16	18	
Arms	Yes	13	23		20	21	
	No	17	27		10	12	
Hips	Yes	14	20		18	16	
	No	16	30		12	17	
Legs	Yes	23	26	4.8[c]*	25	17	7.2[c]**
	No	7	24		5	16	

[a] Yes = change seen; No = change not seen.
[b] Two-tailed test.
[c] One-tailed test.
* $p < .02$.
** $p < .01$.

($\chi^2 = 3.1$, $p < .10$, two-tailed test) than the men. No differences of consequence are present for shoulders, arms, or hips.

Sample 2. The results for Sample 2 even more forcefully indicate that men perceive aniseikonic changes in the leg area with greater frequency than do women. Table 1 indicates that the difference is significant at the .01 level. It is parenthetically interesting that the trend noted in Sample 1 for women to exceed men in frequency of perceived change in the face is clearly supported ($\chi^2 = 5.4$, $p = .02$, two-tailed test). None of the other body sectors were differentiated in more than a chance fashion for the two sexes.

Discussion of Results

The data demonstrate with clarity that men are more receptive than women to the perception of aniseikonic induced perceptual changes in the leg area. It is also noteworthy that the men did not exceed the women in aniseikonic change for any other body region. When one considers that a difference in leg perception was also found by Wittreich and Grace to hold true for children, there can be no question regarding the generality of this phenomenon. It would appear that there is a basic difference between the male and female in our culture in their ability to experience perceptual change in the legs, and it extends into the early years. In terms of the available research regarding the factors underlying aniseikonic change this strongly suggests that the female finds her legs as perceptual targets to be a source of anxiety. She is not sufficiently secure about them to take a chance on their visual transformation. It has already been conjectured that her anxiety stems from the fact that she has learned that mobility and movement in space, which are closely associated with the legs, are unsafe for a female. In that sense her anxiety would be a reflection of the limitations conventionally placed upon the female role.

It was of interest that the only other consistent aniseikonic difference which could be detected in both samples was a trend for women to perceive alterations in the face region more readily than did the men. Parenthetically, one cannot but be struck with the fact that this difference which portrays the women as most receptive to the perceptual alteration, involves a sector which is at the extreme opposite end of the body from the legs. In any case, if one considers that the face is the area most consistently used to register expressive emotion, this finding is congruent with Parsons' (Parsons & Bales, 1955) statement that women are more specialized as expressive rather than instrumental agents in intimate situations (e.g., within the family) than are men. Can it be that the woman is more skilled in the use of her face for expressive communication and therefore more confident and less anxious or defensive about its functioning?

The data with regard to the legs was sufficiently promising to encourage an investigation of whether there are sex differences in degree of activation of the leg region. It is this matter which is explored in the next project to be described.

STUDY 2

Skin Resistance Reactivity Ratios

As previously indicated, there are findings in the literature which suggest that attitudes toward specific body areas may have their physiological correlates. For example, it has been shown in a series of studies that the more definite and clearly articulated an individual's body image boundaries the greater is his tendency to show high reactivity in the boundary regions of the body (viz., skin and muscle) and low reactivity in interior sectors, e.g., heart (Davis, 1960; Fisher, 1963; Fisher & Cleveland, 1958; Shipman et al., in press). Relatedly, ratios of skin resistance between certain body areas have proven to be related to differential values assigned to these areas. A typical pattern has been to find that a body sector given relatively lower prominence than another in the body scheme is also characterized by relatively lower skin activation. Thus, it has been demonstrated that the perception of one's head as small in relation to one's body is correlated with a high skin resistance level on the head (i.e., low activation) as compared to the body. Differences in size ascribed to the back versus front and upper versus lower areas of one's body have been analogously correlated with skin resistance differentials (Fisher, 1958, 1961a, 1961b).

It was this background of work which suggested that if men and women differ in the apparent role they assign to their legs in the body scheme, there should also be a corresponding physiological activation difference. It was reasoned that the female is anxious about the mobility functions of her legs and conflicted about how to integrate them into her body scheme. As symbols of a potentially dangerous kind of activity, she is inclined to minimize their significance in comparison to other body areas. If this be so, it would follow that the degree of activation of the woman's legs should be less in relation to her other body regions than is true for men. It was therefore predicted that the ratio of leg skin resistance to that of an upper nonleg area should demonstrate lower leg activation for the female than the male. Also, since Wittreich and Grace had found aniseikonic leg perception differences in children, it was anticipated that the same activation sex difference would be observed in children.

Procedure

Skin resistance was used as the index of relative physiological reactivity of upper body site in relation to lower body site. It was measured by means of a Brush direct-writing oscillograph. There was a constant current supply of 20 milliamperes and a direct current amplifier for measuring the voltage across subject. Calibration of the record was made directly in ohms. Separate balanced systems were utilized for the upper and the lower site measures. The area of recording from the sites was equalized by means of pieces of tape punched with two holes, each of ¼-inch diameter.

The period of recording was based on the time required for both sites to stabilize to a point of no change for a 15-second period. A minimum of 30 seconds of recording was taken in any case. The median length of recording was 204 seconds. The upper site recording was taken from the dorsal surface of the middle joint of the middle finger of the left hand. A lower site recording was taken from a point ½ inch anterior to the ankle. The choice of the site was quite arbitrary and was based mainly on the ease with which electrodes could be applied. A final reactivity value was tabulated for each subject that was equal to the ratio of the hand resistance level to the lower site resistance level (hand resistance:ankle resistance) at the time of stabilization. The test-retest reliability of this procedure has been shown elsewhere to be adequate (Fisher, 1961a).

Subjects

The subjects consisted of 275 college students (120 male, 155 female). Their median age was 20. They were recruited by payment of a fee.

Included in the study were 230 children (117 boys, 113 girls). Their age range was 7–17, and they were recruited by payment of a fee.

Results

The median skin resistance ratio in the adult male group was 1/4.4 (range 2/1–1/20) and in the female group it was 1/5.5 (range 2/1–1/60). When the scores were trichotomized into as nearly equal thirds as possible, a chi-square analysis indicated that the females had a significantly higher hand to ankle ratio than the males ($\chi^2 = 7.7, df = 2, p < .02$, one-tailed test). That is, the women manifested, as predicted, a relatively lower degree of skin

TABLE 2

CHI-SQUARE ANALYSIS OF ADULT SEX DIFFERENCES IN RATIO OF HAND TO ANKLE SKIN RESISTANCE

Sex	Hand-leg ratio		
	High	Medium	Low
	$\frac{1}{(8.1 \text{ or higher})}$	$\frac{1 \quad 1}{(3.4\text{–}8.0)}$	$\frac{1}{(3.3 \text{ or less})}$
Men	25	49	47
Women	54	58	43

Note.—$\chi^2 = 7.7$, $df = 2$, $p < .02$, one-tailed test.

TABLE 3

MEDIANS AND RANGES OF HAND TO ANKLE SKIN RESISTANCE RATIOS IN BOYS AND GIRLS IN AGE RANGE 7-17

	Age									
	7–8		9–10		11–12		13–14		15–17	
	M (N = 26)	F (N = 19)	M (N = 24)	F (N = 25)	M (N = 22)	F (N = 21)	M (N = 21)	F (N = 22)	M (N = 24)	F (N = 26)
Median	$\frac{1}{10}$	$\frac{1}{8.1}$	$\frac{1}{5.1}$	$\frac{1}{7.1}$	$\frac{1}{5.1}$	$\frac{1}{4.1}$	$\frac{1}{4.1}$	$\frac{1}{4.1}$	$\frac{1}{3.1}$	$\frac{1}{5.1}$
Range	$\frac{2}{1}$–$\frac{1}{25}$	$\frac{1}{1}$–$\frac{1}{21}$	$\frac{1}{1}$–$\frac{1}{20}$	$\frac{1}{1}$–$\frac{1}{22}$	$\frac{2}{1}$–$\frac{1}{19}$	$\frac{3}{1}$–$\frac{1}{18}$	$\frac{2}{1}$–$\frac{1}{15}$	$\frac{1}{1}$–$\frac{1}{19}$	$\frac{3}{1}$–$\frac{1}{13}$	$\frac{2}{1}$–$\frac{1}{20}$

Note.—M = male; F = female.

activation in the leg area (in relation to the hand) than did the men.

The medians and ranges for the children are depicted in Table 3. Because of the limited number of subjects at each level, the data have been analyzed by combining adjoining age groups (viz., 7–8, 9–10, 11–12, 13–14, 15–17). Table 4 indicates that in the 7–8 range significant sex differences do not appear. However, beginning in the 9–10 group the girls have, as predicted, a relatively lower level of leg skin activation than the boys ($\chi^2 = 4.6, p < .02$, one-tailed test). This difference was not evident in the 11–12 or 13–14 groups, but reappeared significantly in the 15–17 category ($\chi^2 = 3.2, p < .05$, one-tailed test). It is parenthetically interesting that children in the 7–8 age group have significantly less leg activation than do children 9–17 or adults.

TABLE 4

CHI-SQUARE ANALYSIS OF SEX DIFFERENCES IN RATIO OF HAND TO ANKLE SKIN RESISTANCE IN AGE RANGE 7-17

Ages	Boys	Girls	χ^2
7–8			
H	13	8	
L	13	11	
9–10			
H	9	17	4.6[a]**
L	15	8	
11–12			
H	11	10	
L	11	11	
13–14			
H	9	10	
L	12	12	
15–17			
H	7	14	3.2[a]*
L	17	12	

Note.—H = above median; L = at median or below.
[a] One-tailed tests used.
* $p < .05$.
** $p < .02$.

Discussion of Results

The skin resistance data fully supported the hypothesis in the adult group, but only in a borderline fashion in the child subjects. Only at ages 9–10 and 15–17 could sex differences in ratio of hand-ankle skin resistance be detected. There is no ready explanation as to why significant differences should appear at these points, but not in the 7–8, 11–12, and 13–14 groups. Of course, the 15–17 group is closest to the adult range, and one is left with the impression that the skin resistance ratio difference is more characteristic of the adult than the child. However, the fact that a difference also appeared as early as age 9 cannot be dismissed. Perhaps there is an early sex difference which is for a period obliterated by other developmental events and then later re-emerges.

The overall results from the aniseikonic lens procedure and the skin resistance measures point to a fundamental sex difference in the manner in which the legs are integrated into the body scheme. In turn, this difference has been speculatively linked with attitudes toward mobility and movement in space. It is striking that a sex difference in leg skin resistance levels was predicted from the aniseikonic data, and that the prediction was generally confirmed. This adds weight to the view presented in detail elsewhere (Fisher & Cleveland, 1958) that body attitudes may play an important role in the degree of physiological activation of various body sectors.

STUDY 3

Body Boundary Differentials

This third study deals with sex differences relating to the individual's perception of his body boundaries. It has become apparent from a number of sources (Fisher & Cleveland, 1958; Schilder, 1935; Wapner, Werner, & Comalli, 1958) that persons differ in their experiencing of the boundary regions of their bodies. Some perceive these regions as definite, well articulated, and clearly separating them from their environs; whereas for others they are vaguely defined and represent only a hazy contour. In the ordinary course of events the existence of a body image boundary is not obvious for most persons seem to know well enough where they end and the outer world begins. However, there are pathological states in which loss of boundary occurs, and some of its functions then become apparent. Thus, brain pathology may so disrupt an individual's boundary awareness that when he is asked with his eyes closed to indicate a spot on his skin where he has been touched he points to a locus in space outside of his body. Wapner et al. (1958) have shown that the more articulated the boundaries of a region of an individual's body the smaller does its size appear to him; and further they have observed that when the boundary is disrupted by ingestion of LSD there is a sense of increased ("spread out") body size. Reitman and Cleveland (1964) have reported that when sensory isolation decreases input into the boundaries they become less articulated, with accompanying changes in the touch threshold of the skin. It has been suggested too (Fisher, 1962) that loss of boundaries underlies the schizophrenic's tendency to confuse his own thought processes with outer reality, as typified in the hallucination. He apparently confuses "inner" and "outer."

Fisher and Cleveland (1958) have devised empirical indexes which are descriptive of the clarity or definiteness of the individual's body image boundary. One of the indexes is referred to as the Barrier score. It is based upon a count of the number of responses to an ink-blot series in which special protective, containing, or decorative functions are assigned to the periphery of the percept. Some examples of Barrier responses follows: cave with rocky walls, woman in a fancy costume, vase, sheepskin coat, turtle with a shell. It has been shown that the Barrier index can be scored with high objectivity; that it has adequate test-retest reliability; and that it is meaningfully linked with reported patterns of body sensation from boundary and nonboundary regions of the body (Fisher, 1963).

A second boundary index devised by Fisher and Cleveland is called the Penetration score. It is based upon a count of all responses to an ink blot series in which the boundary of the percept is somehow open, bypassed, breached, or destroyed. Some examples of penetration responses follow: X ray of body, bleeding wound, vagina, bullet entering flesh, crushed bug. It has been shown that such responses can be scored with high objectivity (Fisher, 1963).

Many studies have been reported which indicate that the Barrier and Penetration scores are predictive of important personality, physiological, and social behavioral patterns. These studies have been reviewed in detail elsewhere (Fisher, 1963). They demonstrate that the more definite an individual's boundaries the more likely he is to perceive himself as possessing clear-cut identity and to have a concept of his body as a well-differentiated sector of space equated with self.

In earlier studies involving the boundary scores it had not been evident that adult sex differences existed (Fisher & Cleveland, 1958). The only previously described sex differences involved children. It was noted that in the age range 5–7 years boys had significantly lower Barrier scores than girls. While in the age range 10–13, girls proved to have lower Barrier scores than boys. The difference observed at the 5–7 level was interpreted as paralleling other findings which pinpoint this period as one in which boys are unusually conflicted about their relationships with father and sensitive to the possibility that he might attack them. Thus, the lowered Barrier score was viewed as reflecting expectation of body damage.

There was no obvious explanation for the sex difference in Barrier scores which was observed for the 10–13 age group, except for the possibility that more girls than boys would in the 10–13 range be already fully involved in the conflicts and difficulties of adolescence. That is, because girls mature faster than boys, it was considered that a higher percentage of them in the 10–13 age class would be coming intensely to grips with adolescent body and role changes and therefore experience an unusual degree of body-focused anxiety (Fisher & Cleveland, 1958).

Although it had been accepted on the basis of earlier studies that there were no adult sex differences in boundary definiteness, hints began to accumulate that small but consistent differences might be present when careful controls were imposed to obtain an equal number of responses from men and women with similar educational and socioeconomic backgrounds. It was the intent of the present study to examine this possibility in a large sample.

Procedure

All Barrier and Penetration scores were based on responses to the Rorschach ink blots which were shown on a screen to subjects in small groups. Response total was controlled by asking that a fixed number of percepts be given to each blot. Three responses were requested for each of cards I, II, III, VII, and X and two for each of the other five cards in the series.

Not only were Barrier and Penetration scores computed, but also an overall index of boundary definiteness which is equal to number of Barrier minus number of Penetration responses. This index has shown itself to be a meaningful measure of boundary definiteness in several studies (e.g., Cassell & Fisher, 1963; Conquest, 1963).

TABLE 5

CHI-SQUARE ANALYSIS OF SEX DIFFERENCES IN BOUNDARY SCORES

Boundary score	Sample		χ^2
	Men	Women	
Barrier			
H	135	184	
L	139	106	11.5[a]*
Penetration			
H	154	118	
L	120	172	13.6*
Barrier minus Penetration			
H	110	182	
L	164	108	28.9*

Note.—H = above median; L = at median or below.

[a] Two-tailed test.

* $p < .001$.

Subjects

A large sample consisting of 564 subjects was assembled (274 men and 290 women). The subjects were all college students recruited either by payment of a fee or as part of a required testing program. The median age in both sex groups was 21.

Results

The median Barrier, Penetration, and Barrier minus Penetration scores in the male group were, respectively, 6 (range 0–16), 3 (range 0–12), and 2 (range from −5 to 13). The respective medians in the female group were 7 (range 1–18), 2 (range 0–12), and 4 (range from −5 to 13).

As shown in Table 5 there is a highly significant trend for males to have lower Barrier scores than females ($p < .001$ level). The men also proved to have higher Penetration scores ($p < .001$ level) and lower Barrier minus Penetration scores ($p < .001$ level) than the women. Thus, in terms of all three indexes women are portrayed as having more definite boundaries than men.

Discussion

It is not a simple matter to account for the greater degree of boundary definiteness displayed by the females. This represents a contradiction of stereotypes which portray the male as superior to the female with regard to such variables as self-definition, clarity of body concept, and body security. A view emphasizing male superiority in this respect was especially fostered by Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954) who found that men were better able than women to utilize body cues in making spatial judgments, and who therefore concluded that men have a more secure and effective body concept. Actually, when one examines Witkin's work, it becomes apparent that his conclusions regarding body concept differences between the sexes are tangential deductions from the fact that women performed more poorly than men in using kinesthetic cues to make spatial judgments when they also have the choice of using visual cues for the same purpose. The fact which is usually overlooked is that when women were placed in situations in which they were forced to rely on their own kinesthetic resources, they proved to be just as accurate in their spatial judgments as men. It is not accurate to interpret Witkin's work as demonstrating that women cannot utilize body cues as efficiently as men. Interestingly, one may note further that when he analyzed sex differences in body concept as defined by figure drawing indexes he was unable to detect any superiority for the male.

Indeed, other studies have provided information which suggests not only that the female devotes more attention to her body than the male, but also that she more quickly arrives at an articulated and realistic concept of it. As already indicated, Katcher and Levin (1955) observed that girls arrive at a realistic concept of their body size at an earlier age than boys. Swenson and Newton (1955) and Wieder and Noller (1950, 1953) reported that girls seem to attain earlier than boys a sexual definition of self as measured by the frequency with which they draw a figure of the same sex as self when asked to draw a picture of a person. They noted also that girls earlier incorporated details into their drawings which clearly distinguished the sex of the figure. Machover (1953) after analyzing the drawings of 1,000 children from ages 5–11 indicated that girls seemed to have less difficulty than boys in evolving a comfortable body concept.

It is possible, then, that the relatively greater boundary definiteness exhibited by the women in the present study represents a clearer articulation of the body concept. Perhaps women exceed men in the clarity with which they perceive their bodies as differentiated and individualized. One could argue that the culture encourages the female to be more interested in her body than it does the man. Also, her role as a woman is more explicitly identified with her body and its functioning than is true of the man. The man's role and status are typically defined in terms of his accomplishments and attainments rather than in terms of his body attributes, but for the woman her role is still largely defined in relation to the attractiveness of her body to the male and her ability to bear children. She learns rather early that her body will be her most important means for the attainment of a meaningful role. It is true that the male is given various messages about the importance of having a strong masculine body, but it becomes evident to him that his success as a man will have little to do in the long run with his body attributes. Only an athlete can see a direct equation between his body strength and success in life. A woman probably more nearly equates self with body. She has a clearer concept than a man of the role her body will play in her life. It is quite apropos to point out further that one of the prime eventual goals of most women involves the conversion of her body into a "container" or protective enclosure for the production of children. Does not the successful conceptualization of one's body as a containing, protective form necessarily mean that it must be experienced as having clear and dependable boundaries?

STUDY 4

Differential Modes of Body Awareness

It has been proposed that individuals differ in general awareness of their own bodies. One may picture each person as distributing his attention between his body and its environs. At one pole are those who are intensely preoccupied with their own body sensations; whereas at the opposite pole are others who seem to have little awareness of their bodies. In a previous study a technique for measuring body awareness (body prominence) was described. It is based on the frequency with which an individual refers to his own body when a sample is taken of what lies within his immediate awareness (Fisher, 1964). A body prominence score derived from such a sample has been shown to have a test-retest reliability of .59 over an intervening period of 3 weeks. It has also proven to be significantly correlated with selective memory for words referring to the body (Fisher, 1964), to be enhanced by conditions which increase body sensations (e.g., ingesting a drug), and to be intensified in situations where nonbody stimulation is decreased (e.g., when isolated by oneself rather than being in a group).[3]

The present project represents an attempt to confirm certain exploratory observations which indicated that men and women differ with regard to the factors that contribute to the prominence of their own bodies in their perceptual fields. The specific hypothesis to be evaluated is as follows. Body awareness in a woman is positively linked with the clarity and definiteness of her body boundaries. However, body awareness in a man is not based upon the definiteness of his body boundaries. Rather, it is positively correlated with the relative salience of the gastrointestinal region in his body scheme. Broadly speaking, this hypothesis suggests that body awareness in women is a function of conditions involving the boundary regions of the body; whereas in men it reflects experiences pertaining to the digestive aspects of the body interior. This would, of course, imply that body awareness has different meaning and consequences in men than it does in women.

The exploratory study which suggested this hypothesis and the cross-validating efforts that followed will be presented in sequence.

Study 4A

Procedure

As part of a larger project, the following three body image measures were administered to the groups of subjects:

1. Body Prominence: The relative prominence of the subject's body in his own perceptual field was measured in terms of what lay within his awareness at a given time. He was asked in a group situation to list on a sheet of paper "ten things that you are aware of or conscious of right now." The 10 responses given were scored by summing the number of references he made to his own body. Such body references were defined so as to include explicit body designations (e.g., "My head hurts"), temperature or kinesthetic sensations, eating experiences (e.g., "I would like to eat a piece of pie"), and descriptions of one's own clothing (e.g., "My dress is green"). Interscorer agreement for two judges for 59 protocols was 95%. The rationale for this measurement is that the greater an individual's perceptual focus upon his own body the more should his body (or appropriate equivalents) find representation in his reports regarding the content of his awareness.

In the present study body-prominence responses were sampled from each subject on three different occasions with a week intervening between the first and second and also the second and third sessions. A total score equal to the sum of the three separate scores which had been obtained was tabulated.

2. Barrier score: This index has been described in detail above in Study 3.

3. Body Focus Questionnaire (BFQ): The BFQ was developed to obtain information regarding the relative prominence of various parts of an individual's body in his body scheme. It attempts to ascertain the distribution of the individual's attention to the major aspects of his body. When BFQ is administered, the subject is presented with a list of 91 paired body areas and given the following instructions:

> Turn your attention upon yourself. Below is a list in which different areas of your body are listed in pairs. In each case circle the area or part which is at the moment most clear in your awareness. In other words, circle the one which at the moment stands out in your attention.

The items included are categorized into nine body dimensions. The greater the frequency with which the items comprising a particular body dimension are selected as "Most clear in awareness" the more perceptually prominent is that dimension considered to be.

In the present instance only those 10 items will be considered that relate to a comparison of one's stomach with a series of noninternal body areas.

---

[3] Author, unpublished data.

TABLE 6

CHI-SQUARE ANALYSIS OF RELATIONSHIP OF BARRIER SCORE TO BODY PROMINENCE IN MALE AND FEMALE GROUPS

		Barrier score					
		Male			Female		
		H	L	χ^2	H	L	χ^2
Body Prominence	H	11	13		16	6	9.5*
	L	14	12		5	15	

Note.—H = above median; L = at median or below.
* $p < .001$.

Subjects

The subjects were 50 male and 42 female college students who were recruited by offering them a fee. The median age in both groups was 20.

Results

The medians and ranges for the Body Prominence, Barrier, and BFQ variables were, respectively, in the male group: 7 (range 1–13), 6 (range 2–13), 5 (range 1–10). In the female group they were, respectively, 7 (range 1–14), 7 (range 2–18), 4 (range 1–10). Relationships between Body Prominence and each of the other two variables were examined by means of the chi square. As shown in Tables 6 and 7, it was found that in the female group Body Prominence was significantly and positively related to the Barrier score ($\chi^2 = 9.5, p < .01$, two-tailed test), but not to the BFQ stomach score ($\chi^2 = 1.9, p < .20$). However, in the male group Body Prominence had only a chance link with the Barrier score and a significant positive one with the BFQ stomach score ($\chi^2 = 5.9$, $p < .02$, two-tailed test). Thus, in the female sample body awareness seemed to be tied in with the degree of definiteness of the body image boundary, but did not relate to the relative position of the stomach in the hierarchy of attention devoted to various body regions. The overall level of body awareness of the males appeared to be a positive correlate of the amount of attention directed to the stomach and was apparently not a function of body boundary conditions.

It was these first exploratory findings which led to the hypothesis regarding difference in the process underlying body awareness in the two sexes.

Study 4B

Procedure

The Barrier and BFQ stomach scores were secured in the same fashion as described in Study 4A. However, the Body Prominence score was obtained by asking the subject to write down 20 rather than 10 things of which he was aware or conscious.

Subjects

The experimental population consisted of 112 women and 109 men who were college students.

TABLE 7

CHI-SQUARE ANALYSIS OF RELATIONSHIP OF BODY PROMINENCE SCORE TO STOMACH AWARENESS IN MALE AND FEMALE GROUPS

		Body Prominence score					
		Male			Female		
		H	L	χ^2	H	L	χ^2
Stomach awareness	H	16	8	5.9^{a*}	14	8	1.9^{a*}
	L	8	17		8	11	

Note.—H = above median; L = at median or below.
[a] Two-tailed test.
* $p < .02$.

The median age in both sex groups was 22. Some of the subjects were recruited by offering them a fee, but the majority was obtained in the course of required testing programs.

Results

The Barrier, Body Prominence, and BFQ medians in the male group were, respectively, 6 (range 0–14), 3 (range 0–10), and 5 (range 1–10). In the female group the respective medians were 7 (range 0–15), 3 (range 0–11), and 5 (range 1–10).

Once again the Body Prominence score was significantly and positively related to the Barrier score in women ($\chi^2 = 4.0$, $p = .02$, one-tailed test), but had only a chance association with the Barrier score in men.

However, while Body Prominence was, as predicted, not related to the BFQ stomach score in women, it also proved, contrary to prediction, to have only a chance relationship to BFQ stomach in men.

Study 4C

Procedure

Only the Body Prominence and BFQ scores were obtained in order to evaluate further the fact that, contrary to prediction, they were not positively related in the male group in Study 4B.

Subjects

The subjects were 34 men and 50 women who were college students. They were tested during regular psychology class sessions. The median age was 20.

Results

The median Body Prominence and BFQ scores in the male group were, respectively, 4 (range 0–8) and 5 (range 0–10). In the female group the respective values were 3 (range 1–11) and 4 (range 0–10).

The data were, in this instance, clearly supportive of the hypothesis. Thus, in the male group Body Prominence and BFQ stomach were correlated .34 (rho) which is significant at the .05 level (one-tailed test). The relationship for the females was of a chance order, as was expected.

Study 4D

Procedure

The Body Prominence, Barrier, and BFQ stomach scores were obtained by the same methods outlined in Study 2.

Subjects

Forty-nine men and 74 women participated. They were college students with a median age of 21. They were primarily recruited by testing scheduled psychology classes.

Results

The Prominence, Barrier, and BFQ scores in the male group were, respectively, as follows: 2 (range 0–9), 5 (range 2–12), and 5 (range 0–10). In the female group they were, respectively, 2 (range 0–10), 6 (range 2–15), and 5 (range 0–10).

Analysis of the data indicated that Body Prominence was positively and significantly related to the Barrier score in the female group ($\chi^2 = 5.6$, $p < .01$, one-tailed test), but not in the male group. As predicted, no relationship was found between Body Prominence and BFQ stomach for the females, but, contrary to prediction, no relationship was also found for the males.

Study 4E

Procedure

The BFQ scores were obtained in the usual fashion. However, when the Body Prominence procedure was administered, the subject was asked to give 10 rather than 20 responses.

Subjects

The subjects were 21 men and 21 women who were relatives of patients hospitalized in a psychiatric hospital. The median age in the male group was 29 years, and the average educational level 10 years. In the female group the average age was 32 years, and the average educational level 11 years.

Results

The median Body Prominence and BFQ scores in the male group were, respectively, 2 (range 0–8) and 5 (range 0–8). In the female group the respective medians were 2 (range 0–7) and 5 (range 0–9).

As predicted, no relation between the Barrier and Body Prominence scores was found in the male sample, but a significant rho of .46 ($p < .05$, one-tailed test) was found in the female group.

Study 4F

The difficulty in obtaining consistent, predicted differences between the sexes with regard to the relationship of the BFQ stomach score to Body Prominence led to the development of another approach to this issue. It was decided to make use of selective memory for words pertaining to a given body region as an index of the prominence of that region in the body scheme. The rationale for this procedure was provided by previous studies in which it was actually shown that there is a positive relationship between the degree to which an individual directs his attention to a given region of his body and the degree to which he demonstrates selectively superior recall for references to that region when he learns a list of body terms. Thus, in one study (Fisher & Fisher, 1964) it was found that the more a subject is aware of the boundary regions (viz., skin and muscle) of his body rather than its interior sectors the more he manifests superior memory for words pertaining to body exterior sensations (e.g., skin itch, muscle ache) as contrasted to words designating interior sensations (e.g., heart beat, stomach ache). Relatedly, it was ascertained in another project (Fisher, 1964) that the greater the attention focused by an individual upon his own body in relation to other objects in his perceptual field the more he selectively recalls words with body meaning rather than words without body implications.

It was anticipated that if the hypothesis were correct that body awareness is linked to gastrointestinal sensation in men but not in women, there should be a significant positive relationship in men between body prominence and selective superior recall for body areas related to gastrointestinal functioning. However, such a relationship should not be found in women.

Procedure

Selective recall for references to gastrointestinal body areas was measured in the following fashion. A list of 20 words pertaining to the body was compiled which follows: SPINE, THROAT, GUT, BRAIN, JOINT, STOMACH, LUNGS, SALIVA, LIVER, LIPS, MOUTH, BLOOD, KIDNEY, TONGUE, HEART, INTESTINE, GUMS, BLADDER, TEETH, SKULL. Ten of the words refer either directly or indirectly to body areas involved in the intake and digestion of food. The other 10 words designate body areas not related to such digestive processes. Both sets of words are of the same average length and randomly assigned to positions in the list. Selective recall was evaluated by telling the subjects (in groups of three to five) that they were to be shown a list of words which they were to study. The list was exposed for 1 minute and removed from view. Then a 5-minute period was provided for the subjects to write on a sheet of paper as many of the words as they could remember. The words recalled were scored by summing those with nutritive references and subtracting the sum of those without such references. Thus, the more positive the score the greater the proportion of nutritive body areas recalled.

Body Prominence was measured in the usual fashion and the subjects were instructed to give 20 responses.

Subjects

Two samples were evaluated. Sample 1 consisted of 58 men and 32 women whose median age was 20. Sample 2 was comprised of 55 men and 28 women. All subjects were college students (median age 20).

Results

In Sample 1 the median Body Prominence score for men was 3 (range 0–10) and the median memory score was 0 (range −5 thru 5). The respective medians in the female group were 3 (range 0–8) and 0 (range −3 thru 4).

The results in the male group are significantly in the expected direction ($\chi^2 = 3.2$, $p < .05$, one-tailed test). The greater the male subject's awareness of his body the more likely he is to show selectively superior recall for body areas involved in the nutritive-digestive process. Also, as predicted, only a chance relationship turned up in the female group between Body Prominence and oral memory.

In Sample 2 the median Body Prominence score in the male group was 3 (range 0–9) and the median oral memory score was 0 (range −4 thru 4). The respective medians in the female group were 3 (range 0–10) and 0 (range −5 thru 5).

The results for Sample 2 were of the

same order as observed in Sample 1. Thus, a rho correlation of .27 ($p < .05$, one-tailed test) was found between Body Prominence and degree of superior recall for the nutritive-digestive words. In the female group the relationship was .12, which is of a chance order.

Discussion of Overall Results

Several studies have been presented dealing with the hypothesis under consideration. That aspect of the hypothesis which states that Body Prominence is positively linked to boundary definiteness in women but not in men is well supported by the data. In four different samples this pattern of interrelationship has been significantly demonstrated. However, it has been more difficult to establish that Body Prominence in men is linked with degree of awareness of nutritive-gastrointestinal areas. While two studies indicated a significant positive link between Body Prominence and BFQ stomach in men, there were also two others in which this could not be shown. When the technique involving selective memory for nutritive-digestive terms was applied, consistent support for the hypothesis did emerge. The more a man is aware of his body the more likely he is to display selectively good recall for nutritive-digestive words. This is not true in women.

What are the prime implications of the findings? They do indicate that intensity of body awareness is founded upon different classes of variables in women and men. In women, awareness of one's body is related to the clarity of one's body boundaries. Whereas in men, a more important contribution to body awareness is made by the nutritive-digestive aspects of one's body scheme. The fact that body awareness in women is a function of boundary definiteness may be most meaningfully interpreted within the context of previous work which had dealt with boundary phenomena in terms of the Barrier score. There are considerable data which portray the person with definite body image boundaries as goal oriented, self-steering, and possessed of an articulated sense of identity. Clear-cut boundaries seem to be accompanied by the ability to behave as a well-individuated person. From this perspective, it would appear that degree of body awareness in women, which is positively related to boundary definiteness, is an expression of individuation and differentiation. The woman who is relatively highly aware of her body may be viewed as one who expresses herself with a clear sense of self-identity. Her body awareness is perhaps one manifestation of her clear-cut differentiation from her environs. Contrastingly, the woman with relatively little body awareness may be characterized as being only a hazily defined individual.

The picture seems to be quite different for men. Body awareness is not related to boundary definiteness, but rather to the prominence of the nutritive-digestive areas in the body scheme. This means that the man who is most aware of his body is also the one who is most focused upon his stomach, gut, and related accessory sectors. Thus, a man's degree of individuation would not appear to have a part in his body awareness. It is difficult to know what specific significance to attach to the role that the gastrointestinal system seems to play in a man's level of body awareness. Most previous observations and speculations concerning gastrointestinal awareness (Blum, 1949; Fenichel, 1945; Freud, 1938) have been inclined to adopt a view which associates such awareness with oral impulses striving for expression or with traits of the "oral character." The man with high body awareness would then presumably be one who is unusually interested in incorporating things and finding adequate sources of "supplies" and gratification. This is an interesting possibility that further studies should be able to test.

It is perhaps not going too far beyond the facts to say that the data indicate that in women degree of body awareness is correlated with indicators of personal maturity (viz., individuality and self-steering). However, in men it is related to variables (i.e., oral preoccupation) which probably have connotations of immaturity. This difference could be speculatively related to the contrasting attitudes which one finds in Western culture toward body awareness in men and women. The socialization of the female child emphasizes a conscious concern with how her body impresses others. She is expected to devote considerable attention to adorning and shaping her body to make it attractive and to communicate her sexual interests. This is a normal and an important aspect of learning the role of a woman. However, the boy, while expected to develop muscular strength and agility, is not encouraged to attend to his body in this self-conscious manner, particularly as a potential object of display and attractiveness. It is interesting in this respect that Van Lennep (1957) has found that typically girls show an increasing concern with body themes in Thematic-Apperception-Test stories as they mature beyond adolescence, whereas boys manifest a decreasing degree of such concern. Relatedly, Harlow (1951) reported that men who devoted special attention to building up their bodies by means of weight lifting were likely to be insecure and in considerable conflict about being heterosexually expressive.

STUDY 5

Response to Body Distortion

The findings from the boundary and body prominence studies implied that women have less difficulty than men in dealing with the whole problem of body awareness and arriving at a view of their bodies as distinct articulated entities. There was the implication that women have body concepts which are more closely linked to their identities and therefore more meaningful to them. Within this framework, two hypotheses were formulated which propose relatively greater stability for the female body image.

It was first hypothesized that women have a clearer concept than men of the sexual identity of their bodies. That is, they are presumed to have a clearer concept of their body attributes as congruent with a specific life role. If so, it should follow that they would be less confused or disturbed than men when perceiving themselves in a situation designed to introduce ambiguity concerning the sexual attributes of their appearance.

Secondly, it was conjectured that women are more accepting than men of the immediate structure and configuration of their bodies. Therefore, it was logical to predict that they would be less likely than men to perceive gross alterations in their apparent body proportions as having positive or enhancing value. That is, women were presumed to have less need to seize upon arbitrary externally imposed changes in their body configurations as representing a shift toward some idealized standard.

Study 5A

This study was concerned with the first hypothesis regarding male-female differences in responses to alteration in the sex of one's appearance. It was considered that the more insecure an individual is about the sexual identity of his body the greater will be his anxiety and disturbance when confronted with images of himself that are ambiguous with regard to sexual categorization. If men are more insecure than women about their sexual identity, they should be characterized by signs of relatively greater anxiety when confronted by a situation which introduces contradictions regarding the sex of one's appearance.

Procedure

In order to test this formulation, each subject was asked to describe his own mirror image while wearing a series of male and female masks. The subject was brought into a totally dark room and placed 4 feet from a full-length mirror. He was told to keep his eyes closed until a signal was given. He was further told that a series of masks would be placed upon his face and that each time he was asked to open his eyes he would have a very brief glimpse of his masked appearance in a mirror. His task was to examine his image until it was no longer visible, then he was to close his eyes and describe in as much detail as possible the appearance of his masked face. The subject was given a verbal preparatory signal ("Get ready") 1 second before he was to open his eyes and then two lights (75 watt) were turned on for 1 second by a timer. Simultaneously with the turning on of the illumination the subject was given a signal to open his eyes. Eight different rubber masks were placed upon the subject's face. Four were male and four female and were presented in random order. They were chosen so that their sex characteristics were clearly evident. When the subject described the appearance of his masked face, his words were written down as verbatim as possible.

Scoring of the subject's descriptions was entirely concerned with the correctness of his identification of the sex of each mask. The total number of misidentifications was tabulated. This total could range theoretically from 0 to 8.

It was presumed that confrontation with a situation in which the sex of one's appearance could unpredictably vary would create a threat to the subject in proportion to his own insecurity about his sex attributes. Further, it was expected that the more intense the threat to the subject the greater the likelihood that he would experience anxiety of sufficient magnitude to interfere with the accurate perception of the sex characteristics of the masks. In terms of the hypothesis, this meant that men should make more errors than women in identifying the sex attributes of the masks.

Subjects

Two separate samples of subjects were studied. Sample 1 consisted of 63 subjects (30 male, 33 female), and Sample 2 was comprised of 78 subjects (29 male, 49 female). All were college students recruited by payment of a fee. Median age in both samples was 20.

Results

Sample 1. The median number of sex misidentifications in the male group was 3 (range 0–4) and in the female group 1 (range 0–2). As shown in Table 8, a chi-square analysis confirmed the prediction that the male subjects would report a larger number of sex misidentifications than the female subjects ($\chi^2 = 25.2$, $p < .001$).

Sample 2. The median for sex misidentifications in the male group was 2 (range 1–3) and in the female group 1 (range 0–4). When the data in this cross-validation sample were examined by means of the chi square sex misidentifications proved to be significantly more frequent in the male than female groups ($\chi^2 = 7.4$, $p < .001$).

It should be noted that an analysis of the specific types of misidentification errors in Samples 1 and 2 did not reveal any significant trends for these errors to cluster on masks of the same sex or opposite sex as the subjects.

Discussion

The differentiation between the men and women in number of sex misidentification errors was surprisingly sharp in two different samples. Men seem less able than women to make valid discriminations between male and female sex attributes under the conditions set up in the present experiment. It must, of course, be acknowledged that the fact that the subjects responded to masks on their own faces does not insure that they identified them with their own appearance. The same results could perhaps have been obtained from reactions to the masks had they not been on the subject's own face. However, the fact must be considered that although a multitude of studies have been done of sex differences in responding to many kinds of pictures and blots, none have ever indicated that men are less successful than women in making valid distinctions between that which is male and female. In any case, whether the results of the present study could be duplicated with masks not worn by the subject the basic fact would

TABLE 8

CHI-SQUARE ANALYSIS OF SEX DIFFERENCES IN MISIDENTIFICATIONS OF MASKS IN SAMPLES 1 AND 2

		Sample 1			Sample 2		
		Male	Female	χ^2	Male	Female	χ^2
Sex misidentifications	H	27	9	25.2*	18	15	7.4*
	L	3	24		11	34	

Note.—H = above median; L = at median or below.
* $p < .001$.

remain that men evidence more disturbance than women in making the male-female discriminations required. One should also note that the relatively greater number of errors made by the men in this unstructured situation is opposite to the usual finding that women make more errors than men in vaguely defined judgmental situations (e.g.,Witkin et al., 1954).

Study 5B

The hypothesis was examined that women are more accepting than men of their immediate body attributes and therefore less likely to interpret gross alterations in their appearance as representing "improvement" in the direction of some idealized standard. Stated in another way, it was expected that men would exceed women in reading desirable implications into visual transformations of their bodies. This hypothesis was derived, first of all, from the already stated view that a woman feels more secure about, and basically accepting of, her body than a man because she can more clearly relate its attributes to her primary social role. Secondly, it was considered that women devote a good deal of energy to experimenting with cosmetics and various types of clothing in an effort to arrive at a "best appearance," whereas men are restricted by custom to a minimum of such experimentation. Thus, women have a greater opportunity than men to "act out" their fantasies about how they would like to look to others. This would provide them with a more realistic experiential basis for deciding what improves their appearance. Also, it would mean that they would have fewer unexpressed impulses relating to the alteration of their appearance, and therefore would have less need to capitalize on wishful opportunities for transformation of their bodies.

Procedure

In order to provide subjects with an opportunity to react to gross alterations in their appearance they were asked to stand 4 feet from a full-length mirror and to view themselves while wearing lenses that changed their apparent proportions. One set of lenses was designed to produce obvious lengthening of the vertical axis of the body. They consisted of a pair of equivalent meridional afocal iseikonic lenses of 14% magnification (right eye and left eye—14% [M = 1.20] meridional afocal lenses at axis 180°). Their overall effect was to make the subject look taller and thinner. A second set of lenses was used which produced obvious widening in the horizontal axis. They were comprised of a pair of equivalent meridional afocal iseikonic lenses, of 14% magnification (right eye and left eye—14% [M = 1.20] meridional afocal lenses at axis 90°). Generally, they confronted the subject with an image of himself which was shorter and wider than usual.

The subject was instructed to observe his mirror image and to describe the ways in which it appeared to be altered. Upon completion of his description which was written down as verbatim as possible, he was asked to imagine the kind of person his altered appearance made him resemble. Responses were obtained first for the vertically and then for the horizontally distorting lenses.

The protocols were scored by rating the subject's description of his altered appearance as favorable or not favorable. Two judges who had no knowledge of the hypothesis under consideration attained 86% agreement in their classification of the responses for the total sample. Disagreements were later resolved by joint discussion.

Examples of responses judged as favorable are as follows: Abraham Lincoln, movie star, agile, athletic, like a model, tall and slim. The following are illustrative of responses judged as unfavorable: lazy, fat, clown, weak, old maid.

Subjects

There were 81 subjects (31 men, 50 women). They were college students whose median age was 20 years.

Results

Without exception, all subjects detected the lengthening impact of the vertical lenses and the widening effect of the horizontal lenses. However, there were many individual differences in the interpretations that were placed upon these changes. Some subjects greeted them as highly desirable ("I wish I looked like that"), whereas others rejected them as "ugly" and alien.

As can be seen in Table 9, the men were significantly more often approving than the women of the changes they observed in their mirror images. The differences were highly significant for both the vertical ($p < .001$) and horizontal ($p < .001$) conditions. It was the men who seemed to be intrigued with the potential improve-

TABLE 9

CHI-SQUARE ANALYSIS OF SEX DIFFERENCES IN RESPONSE TO VERTICAL AND HORIZONTAL LENS-INDUCED CHANGES IN MIRROR IMAGE

	Vertical			Horizontal		
	Male	Female	χ^2	Male	Female	χ^2
Favorable	19	6[a]	20.8*	17[b]	11	11.7*
Other	12	42		11	39	

[a] Only 48 female subjects are reported because 2 did not participate in the vertical distortion task.
[b] Only 28 male subjects are reported because 3 did not participate in the horizontal distortion task.
* $p < .001$.

ment in body form suggested by the novel mirror image. Typically, this potential improvement was defined in terms of greater strength or forcefulness.

Discussion

The sex misidentification data and the results from the lens distortion procedures were congruent with the view that women have a more clearly articulated body concept than men. The sex misidentification data are rather unequivocal. However, one could debate an alternative explanation for the lens findings. Is it possible that the women, in rejecting the lens changes, were actually evidencing anxiety about perceiving gross alterations in their appearances? Perhaps the more favorable attitude of the men toward such alterations was indicative of their ease in accommodating to them? These possibilities cannot be dismissed, but they are contradicted by the fact that women are generally more daring and less embarrassed than men in exhibiting themselves publically in novel clothes styles. It should also be pointed out that the favorable reactions of the men to their altered images occurred whether these alterations involved becoming taller and thinner or shorter and wider. That is, even though the two classes of alteration were directly opposite in character they both evoked favorable reactions from the men. This appears to be an expression of a diffuse set rather than of a discriminating openness to new body experiences.

GENERAL CONCLUSIONS

A number of categories of sex differentiation in body concept have been explored. It was intended to demonstrate that in addition to sex differences involving the so-called phallic-nonphallic dimension which has been described there are other differences that are perhaps linked with the cultural roles assigned to the male and female. Evidence has been presented that women have more anxiety and uncertainty about their legs than men, and it was successfully predicted from this fact that their legs would be relatively less activated (in terms of skin resistance level) than those of men. The theory was offered that such leg differences in the body scheme are a function of the fact that women are made to feel that mobility and movement, which are of course closely associated with one's legs, have potentially dangerous consequences.

Data were also presented that indicate a small but consistent tendency for women to experience their body boundaries as more definite and articulated than do men. A possible explanation for this finding was formulated in terms of the idea that the culture spells out clearly for the woman how her body is related to the fulfillment of her principal life goals, whereas the man finds it difficult to associate his body attributes with what is required of him for success and attainment. That is, women have clearer criteria for conceptualizing their bodies as psychological objects. Additional support for this view was offered by findings which demonstrated that men are more disturbed than women when perceiving their mirror images in a situation which introduces conflict as to the sex attributes of their appearance. In still another study it was shown that whereas degree of body awareness in women is linked with the definiteness of the body image boundary, it seems in men to reflect the prominence of gastrointestinal experiences. This implies that body awareness is a more mature manifestation in the female than in the male. Perhaps this is a function of the fact that it is an expected part of the female's role to focus attention upon her body. For a woman not to be "body aware" is to signal lack of femininity. By contrast, men are expected to be muted in the display and awareness of their bodies. Current standards of mature and successful male behavior do not assign much significance to body awareness, as such.

The studies just reviewed suggest that there exists a spectrum of sex differences in body concept which is related not only to obvious matters of anatomy but also to particular cultural definitions of masculinity and femininity. It is likely that each culture's arbitrary concepts of masculinity-femininity may result in specific sex differences in body image. Thus, in a culture which emphasized female mobility and male passivity there might be a reversal of the present findings which indicate that anxiety about the leg region is greater in women than in men. Or in a setting where masculinity was primarily expressed via display of body strength and agility (e.g., as might be true of a warrior) body awareness in the male might prove not to be related to gastrointestinal sensations but rather to the degree of articulation of his body boundary.

This perspective which relates sex differences in body concept to the values of the culture is a logical derivative of a position outlined elsewhere (Fisher & Cleveland, 1958) which proposes that an individual's body attitudes reflect in body terms the important values and identifications he has evolved. From this view a man's concept of his body size might be significantly influenced by the degree to which he is identified with an inferior depreciated role. Or his degree of awareness of his mouth might be an expression of attitudes he has learned about incorporating. Quite analogously, members of each sex may adopt specific attitudes toward their bodies which mirror the division of function and status for the sexes typifying their particular culture.

The point should be made that the overall data do not demonstrate a generalized superiority in body concept for men. There has been a tendency in terms of Freud's speculations and Witkin's observations to assume that men display better integration and more effective organization of the body image than do women. However, the present results raise the opposite possibility that in many ways women have less difficulty than men in making psychological sense of their bodies.

It is not farfetched to speculate that the sex differences in body concept which have been discovered may provide useful leads and methods for studying sex-role problems within each sex. There has been little success in finding dependable techniques for measuring degree of masculinity or femininity. It is possible that measures based on body feelings and attitudes are particularly sensitive to the masculinity-femininity dimension. Thus, an evaluation of how an individual regards certain areas of his body or his reactions to the perceptual alteration of his body may prove to be sensitive indicators of the kind of sex role he has learned.

SUMMARY

A series of studies has been presented designed to examine sex differences in body perception and body concept not directly related to phallic-nonphallic distinctions. The following were the principal findings:

1. Men proved to be more receptive than women to the perception of aniseikonically induced alterations in the appearance of their legs. This was supportive of findings reported by Wittreich and Grace and may be interpreted as congruent with the view that women have relatively high anxiety about their legs as a consequence of the negative view taken by the culture toward mobility in space by the female.

2. The demonstration of such a sex difference in aniseikonic perception of

one's legs suggested that women are inclined to minimize the prominence of their legs in the body scheme. This, in turn, resulted in the prediction that females would be characterized by relatively less physiological activation of the legs (in relation to an upper body sector) than would men. The hypothesis was confirmed in adult subjects, but only partially supported in a group of children.

3. Assumptions regarding the superiority of the male to the female in body concept were contradicted by the finding that women obtain significantly higher Barrier, lower Penetration, and higher Barrier minus Penetration scores than men. Women appear to have a more distinct concept of their body boundaries than men.

4. It was further shown that for women the prominence of one's own body in the perceptual field is significantly and positively linked with boundary definiteness, whereas for men it proved to be positively related to degree of awareness of gastrointestinal sensations. These findings indicate that body awareness is a resultant of different variables in men and women and also imply that it is more likely to be associated with a clearly articulated identity in the latter.

5. The fact that the data did not seem to support the usual suppositions about men having a more clearly defined and stable body concept than women stimulated two other studies of the issue. One demonstrated in two different samples of subjects that men make more errors than women in identifying the sexes of masks worn on their own faces in an ambiguous setting. A second indicated, as predicted, that men were more likely than women to interpret lens-induced changes in the vertical and horizontal axes of their mirror images as improvements upon their actual appearance. The data from these two studies seem to portray women as having a more individuated and meaningful body concept than men.

The overall results indicate that males and females do differ with regard to a number of body-image dimensions which are not directly related to the phallic-nonphallic distinction. In addition, they provide converging lines of evidence which suggest that the female may have a more definite and stable concept of her body than the male.

REFERENCES

BENNETT, D. H. The body concept. *J. ment. Sci.*, 1960, **106**, 56–75.

BLUM, G. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, **39**, 3–99.

CALDEN, G., LUNDY, R. M., & SCHLAFER, R. J. Sex differences in body concepts. *J. consult. Psychol.*, 1959, **23**, 378.

CASSELL, W. A., & FISHER, S. Body image boundaries and histamine flare reaction. *Psychosom. Med.*, 1963, **25**, 344–350.

CONQUEST, R. A. An investigation of body image variables in patients with the diagnosis of schizophrenic reaction. Unpublished doctoral dissertation, Western Reserve University, 1963.

DAVIS, A. D. Some physiological correlates of Rorschach body image productions. *J. abnorm. soc. Psychol.*, 1960, **60**, 432–436.

ERIKSEN, E. H. Sex differences in the play configurations of preadolescents. *Amer. J. Orthopsychiat.*, 1951, **21**, 667–692.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FISHER, S. Body image and asymmetry of body reactivity. *J. abnorm. soc. Psychol.*, 1958, **57**, 292–298.

FISHER, S. Body image and upper in relation to lower body sector reactivity. *Psychosom. Med.*, 1961, **23**, 400–402. (a)

FISHER, S. Front-back differentiations in body image and body reactivity. *J. gen. Psychol.*, 1961, **64**, 373–379. (b)

FISHER, S. Body image boundaries and hallucinations. In J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962.

FISHER, S. A further appraisal of the body boundary concept. *J. consult. Psychol.*, 1963, **27**, 62–74.

FISHER, S. Body awareness and relative memory for body versus non-body references. *J. Pers.*, 1964, **32**, 138–144.

FISHER, S., & CLEVELAND, S. E. *Body image and* *personality*. Princeton, N. J.: Van Nostrand, 1958.

FISHER, S., & FISHER, RHODA. Body image boundaries and patterns of body perception. *J. abnorm. soc. Psychol.*, 1964, **68**, 255–262.

FRANCK, K., & ROSEN, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, **13**, 247–256.

FREUD, S. Three contributions to the theory of sex. In, *The basic writings of Sigmund Freud*. New York: Random House, 1938.

FREUD, S. Some psychological consequences of the anatomical distinction between the sexes. In, *Collected papers*. New York: Basic Books, 1959.

HARLOW, R. G. Masculine inadequacy and compensatory development of physique. *J. Pers.*, 1951, **19**, 312–323.

JOURARD, S. M., & SECORD, P. F. Body cathexis and the ideal female figure. *J. abnorm. soc. Psychol.*, 1955, **50**, 243–246.

KATCHER, A., & LEVIN, M. M. Children's conceptions of body size. *Child. Develpm.*, 1955, **26**, 103–110.

MACHOVER, K. Human figure drawings of children. *J. proj. Tech.*, 1953, **17**, 85–91.

PARSONS, T., & BALES, R. F. *Family socialization and interaction process*. Glencoe, Ill.: Free Press, 1955.

REITMAN, E. E., & CLEVELAND, S. E. Changes in body image following sensory deprivation in schizophrenic and control groups. *J. abnorm. soc. Psychol.*, 1964, **68**, 168–176.

SCHILDER, P. *The image and appearance of the human body*. London: Kegan Paul, Trench, Trubner, 1935.

SECORD, P. F. Objectification of word-association procedures by the use of homonyms: A measure of body cathexis. *J. Pers.*, 1953, **21**, 479–495.

SHIPMAN, W. G., OKEN, D., GRINKER, R. R., GOLDSTEIN, I. B., & HEATH, H. A. A study in the psychophysiology of muscle tension. II: Emotional factors. *Arch. gen. Psychiat.*, in press.

SWENSON, C. H., & NEWTON, K. R. The development of sexual differentiation on the Draw-A-Person Test. *J. clin. Psychol.*, 1955, **11**, 417–419.

VAN LENNEP, D. J. Projection and personality. In H. P. David, & E. Von Bracken (Eds.), *Perspectives in personality theory*. New York: Basic Books, 1957.

WAPNER, S., WERNER, H., & COMALLI, P. E., JR., Effect of enhancement of head boundary on head size and shape. *Percept. mot. Skills*, 1958, **8**, 319–325.

WEINBERG, J. R. A further investigation of body cathexis and the self. *J. consult. Psychol.*, 1960, **24**, 277.

WIEDER, A., & NOLLER, P. A. Objective studies of children's drawings of human figures. I: Sex awareness and socioeconomic level. *J. clin. Psychol.*, 1950, **6**, 319–325.

WIEDER, A., & NOLLER, P. A. Objective studies of children's drawings of human figures. II: Sex, age, and IQ. *J. clin. Psychol.*, 1953, **9**, 20–23.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

WITTREICH, W. J., & GRACE, M. Body image and development. Technical Report, March 29, 1955, Princeton University, Contract N6 ONR-270(14), Office of Naval Research.

WITTREICH, W. J., & RADCLIFFE, K. B., JR. The influence of simulated mutilation upon the perception of the human figure. *J. abnorm. soc. Psychol.*, 1955, **51**, 493–495.

(Received April 30, 1964)

Vol. 78, No. 15 1964, Whole No. 592

Psychological Monographs: General and Applied

VARIETIES OF PROJECTION IN TRAIT ATTRIBUTION[1]

DONALD T. CAMPBELL, NORMAN MILLER,[2] JACOB LUBETSKY,[3] AND EDWARD J. O'CONNELL[4]

Northwestern University

Personality has been assessed by peer reputation, and projection has been assessed by the attribution of traits to peers and to unknown photographed persons, using data from 214 men and 236 women. No evidence of similarity projection has been found even when insight, measured in 3 different ways, has been taken into consideration. Contrast, predicted from adaptation-level theory, has been found for the traits "gullible," "secretive," "orderly," "influential," and "suspicious." The predicted instances of complementary projection (a form of projection following from cognition and learning theories) were not confirmed in general. Whereas complementary projection and reactivity processes could explain most of the significant off-diagonal projection correlations (i.e., correlations in which the personality trait and the projected trait differ), the preferred interpretation is to regard these also as showing a contrast process in which assertive and exploitative persons judge others to be passive and gullible.

THIS study was undertaken both because of the importance of the concept of projection in personality theory and because learning and cognition theories also predict such distortions in judgment. Projection thus could be a central concept in psychology, potentially capable of joining our laboratory-developed theories with the fascinating considerations of human personality which continue to be the fountainhead of interest in our field. This centrality is felt to justify both the presentation of our results even when these are disappointing and the discussion of theoretical considerations even when the data presented do not increase their credibility.

In the setting of this research, projection is measured by a correlation between a person's reputed personality traits and the idiosyncrasies in the traits he attributes to others. Theories of acquired behavioral dispositions, as represented in social cognition, learning, and perception, predict a variety of types of projection for this setting. In general, they favor a *complementary* relationship rather than a *similarity* one, although panglossian projection and cassandran projection are also possibilities. In completing the consideration of processes that would generate a reputation-attribution correlation, good theoretical cases can also be made for *contrast* and *reactivity*. The dispositional theories leading to these diverse and often contradictory predictions will be discussed first, followed by a presentation of the research methods and results.

EARLY MEANINGS OF PROJECTION AND THEIR SIGNIFICANCE

Psychological meanings for the noun *projection* and the verb *to project* were already a part of the common vocabulary of European scholars when Freud first used them. It seems clear that both primary and secondary meanings evolved from the study of *optics*, the transfer and reproduction of images through the projection of light rays. This is the origin of "projective geometry," and of the "projector" as lens or other device for throwing an image on a screen. The camera obscura, the oldest of such devices, was a major intellectual preoccupation for as varied a range of scholars as Roger Bacon, Francis Bacon, Da

---

[1] This research was supported by Research Grant M-1544 from the National Institutes of Health.

[2] Now at Yale University.

[3] Now at Middletown State Hospital, Connecticut.

[4] Now at Carnegie Institute of Technology.

Vinci, Kepler, and Newton. Perhaps because the study of optics was central to scientific inquiry, optical phenomena became prominent as a source of analogies for the expression of other ideas. Such analogizing is evident in Francis Bacon, in 1620, who discussed the psychology of *bias*, the false idols or images of the mind:

> For the mind, darkened by its covering the body, is far from being a flat, equal, and clear mirror that receives and reflects the rays without mixture, but is rather an uneven mirror which imparts its own properties to different objects and distorts and disfigures them [Bacon, 1853, pp. 207, 390].

Within this same family of analogies, the words *to project* and *projection* came to mean the external (objective) referral of central-nervous-system events in perception, with or without veridicality in such reference. James (1890, Vol. 2, pp. 31–35, 228–231) gives us a good sample of such usage, citing, among others, Bain, 1855; Taine, 1870; and Schopenhauer, 1847. He uses the verb *to project* incidentally in a discussion of hallucination (James, 1890, Vol. 2, p. 115). Sully (1881, pp. 63, 111, 113) also gives several usages in which the source of central-nervous-system events is mistakenly ascribed to an external reality which in fact does not exist. It is clear that the modern connotation of misperception was already present.

The *Oxford English Dictionary* defines *projection* (in Meaning 9) as

> a mental figure or image visualized and regarded as an objective reality.

Project (in Meaning 10, as a transitive verb) is not defined, but is illustrated by these two pre-Freudian uses: from Robertson's *Sermons* of 1850,

> Then we project everything stamped with the image of our own feelings;

and from Harlan's *Eyesight* of 1879,

> An excited or disordered brain may project some phantasm of its own conjuring . . . and see it as distinctly as if it were a tangible object.

In the 1954 *Merriam-Webster Dictionary* the bulk of the psychological meanings have this pseudoperceptual connotation: *Project* (vt; 5)

> to externalize and regard as objective or outside of oneself, as a sensation, image, or desire.

Projection (6)

> Act of externalizing or of perceiving a mental object as spatially and sensibly objective, also, the object projected.

(12)

> The act of externalizing or objectifying what is primarily subjective.

Only in 12b do we get the limited "similarity" notion.

> The ascribing of one's own thoughts, wishes, inhibitions, and faults to objects or persons in the environment.

This limitation to a similarity usage is not implied even in Rickman's (1937) effort to provide an orthodox psychoanalytic definition of *projection:*

> ascribing mental processes that are not recognized to be of personal origin to persons or situations in the outer world.

The visual afterimage may be taken as a prototype for the pseudoperceptual use of the term projection. Because of an afterimage, one may "project" upon a blank wall an image which is not objectively there. The visual afterimage prototype and the older uses illustrate three important aspects of projection. (*a*) There is an external ascription of the image. The projection is not recognized as such by the projector. He is unaware that he contributes this content. He projects in the course of trying to describe external reality. From this comes the importance in projection research of a phenomenally objective task (Campbell, 1950, 1957; Campbell & Burwen, 1956). The judgment of peers and of photographs, as presented to respondents in the present research, represents such an objective task assignment, while many projective tests do not. (*b*) The projected content correlates with some aspect of the projector's person, be it transient satiation patterns on his retina or more permanent aspects of his personality. (*c*) In most cases of projection in which we are interested, the aspect of personality correlated with the projection is a product of past experience, and through this the projected content is correlated with past external realities. The afterimage which one now projects upon an objectively homogeneous grey wall is a function of a real external stimulus pattern which was a significant part of one's past experience. Of course, in fantasy expression past experience is often elaborated, distorted, or exaggerated; but the emphasis here is on the relation between the contemporary perception and a stored veridical representation of past experience. In other words, though typically defined as a correlation between personality and misperception, projection also implies a correlation between misperceptions of the present external objects and veridical perception of significant external objects of the past. Thus it corresponds to transfer and generalization processes in learning studies and to the transference phenomena of the psychotherapeutic situation.

TYPES OF PROJECTION REFLECTING DISPOSITIONAL TRANSFER

Complementary Projection

We have in psychology and in other social sciences, a great number of concepts referring to the residues of past experience as retained by the individual and as influencing his perceptions and actions. The present project stems from an attempt to catalogue and integrate these concepts (Campbell, 1963) with particular emphasis upon resolving the conflict between cognitive and behavioristic terminologies. The assumption is made that all of our dispositional theories make relevant predictions for the problem setting of any one dispositional construct. Point *c*, above, seems to place the concept of projection among such acquired dispositional constructs. The dispositional theories of social perception and of learning ought, therefore, to make relevant predictions about projection. They should specify particular relationships between the personality traits of the perceiver and his misperceptions of other persons. It turns out that applied at their simplest level, both social cognition and learning theories predict the same general relationship—the attribution of traits or properties to objects which justify or "explain" one's actual behavior toward them. We label this type of relation between a person's personality and his description of the social world complementary projection. Let us consider in order its derivation from social cognition theory and learning theory.

Complementary Projection and Social Cognition Theory. Most, if not all, of the contents of consciousness are "perceptual" or "quasi-perceptual" in nature, portraying external objects or events, and occurring in the mode of visual or auditory imagery. If we can judge from the studies of cortical excitation (e.g., Penfield & Rasmussen, 1950), conscious experience is a concomitant of neural activity in many parts of the brain. In the sensory projection area raw sensations, meaningless flashes of light, fragments of sound, and fragmentary tactile experiences may be activated. *Meaningful* conscious experiences come from other parts of the brain and represent memory or the momentary sensory input from the contemporary external environment filtered through the association areas and thus mixed with engrams from the past. Such conscious experiences are quasi-perceptual in nature; and it is about these, or in analogy with these, that the growing consensus of social-perceptual or social-cognitive theory has emerged. Perception in this sense refers to an image of a present object which has been encrusted, enhanced, or given meaning by the learned values and connotations associated with it. The neural input from the sense receptors has been filtered through the association areas by the time such conscious experience takes place. The learning associated with the external stimulus is thus embodied in the person's perception of it. His way of viewing an object contains or implies his learned ways of responding to it.

The typical program of this perceptual

or cognitive social psychology is to diagnose the person's phenomenal field (by interviewing him for example) and to relate these conscious contents to behavior, action, locomotion, etc. Usually it is not the phenomenal field of the moment, but more enduring idiosyncrasies in viewing the world, of valuing objects, etc., which are involved. From several theoretical efforts of this type comes a remarkable uniformity of predictive rule. According to Lewin (1936, 1951) behavior is a function of the "life space," which can be regarded as a codified "perception" of the relevant aspects of the world. Murphy (1947) has spoken of the "unity of perception and action"; Sherif and Sherif (1956) of the "unity of experience and action," connoting by *experience* conscious awareness. Asch (1952, p. 159) has used the dramatically strong phrasing "isomorphism of experience and action." While he uses the concept explicitly in a very limited setting, it seems convenient to extend it to his more general point of view, as expressed in this quotation,

> We act and choose on the basis of what we see, feel, and believe.... When we are mistaken about things we act in terms of our erroneous notions, not in terms of things as they are [pp. 64–65].

For all of these theories, behavior is correlated with conscious experience in a one-to-one fashion. The perception or the experience contains or already has represented in it the learned determinants of behavior. Consider a schematic neurology in which the brain circuits are strung out in a linear message flow having as its major stages receptors sensory projection areas, association tracts, and motor nerves. At what stage in this series is conscious experience a concomitant of nervous activity? The principle of isomorphism of experience and action seems to say that the conscious experience social psychologists are interested in occurs at the very terminus of the association areas, with no further association tracts intervening to reduce the perfect correlation between experience and action.

These social-perceptual concepts are close to the older notions of projection. To apply them to the correlation between a person's personality traits and the traits he exaggeratedly perceives in others, it is necessary to interpret a person's traits as corresponding to his actions. In our research setting, at least, the trait ratings which a person *receives* must reflect his behavior as observed by others. The social-perceptual prediction is that his behavior corresponds to his perception of the world. The traits he *ascribes* to other persons represent his perception or experience of an important segment of that world. Thus, if a person mistakenly perceives that he is surrounded by hostile persons, he should respond accordingly, perhaps behaving with an anxious or retaliatory hostility and thereby gaining the reputation of a hostile. If he mistakenly perceives everyone to be friendly, he should behave in a way appropriate to that perception, responding in a confident, friendly fashion. Taking the social-perceptual orientation to heart, we should ask ourselves whenever we meet a person with a strongly idiosyncratic way of behaving: What kind of a mistaken view of the world, what kind of a phenomenal field, would lead him to behave in this way? Does this rebellious and touchy person see others as stuffed-shirts trying to dominate him? Does this generously bossy person see others as incompetent and in need of some sound advice? Does this secretive individual see all others as interested in prying into his personal affairs? Does this stingy person see others as greedily trying to take his goodies away from him? The theory has not yet spelled out the correspondence between perception and action in this detail. Lewin, in his life-space diagram, provided only two kinds of significant perceptions (positive and negative valence) related to two kinds of actions (locomotion toward and away from). From these we can at least predict that persons with approaching tendencies will perceive others as of positive valence; persons with avoidant tendencies will perceive others as having negative valence. However, when this same general orientation is combined with folk wisdom, we can fill in many instances of more specific predictions. Some of these are presented in Table 1 under the column headed Complementarity. (Other columns of Table 1 present predictions derived from alternative mechanisms which will be discussed below.)

The prediction from social perception theory thus turns out to be equivalent to what Murray (1933) called "complementary apperceptive projection" in psychology's first explicit use of a projective test. In this classic study, "The Effect of Fear on the Estimates of the Maliciousness of Other Personalities," Murray found that a frightening game of "murder" increased girls' tendency to judge unknown photo persons as malicious, rather than increasing their tendency to see the people as friendly and protective, as wishful thinking (here called panglossian projection) would have made them do.

While in neither source of the theory is temporal priority or causality necessarily invoked, there does seem to be an implicit difference between these illustrations of them. The social-perceptual orientation sees the misperception as causing the characteristic behavior. Note in particular the quotation from Asch. Murray, however, implies the opposite relation: once the general mood or reaction tendency is established, the person then perceives stimuli to be of such a nature as to justify the mood. Whenever the projected content is interpreted as rationalizing or justifying the person's response tendencies, complementary projection as described by Murray is imputed. In Freudian discussions of projections this is clearly a frequent interpretation, though not the only interpretation. Thus, the person with an amorphous guilt and anxiety, produced by fantasied or real actions long since repressed, will perceive newly met persons in his present environment as accusing him of sins. The own trait of anxiety and the projected trait of accusation stand in a complementary relationship to each other. The projected content rationalizes and explains away the anxious foreboding, rather than causing it or being antecedent to it. For our present purposes, however, either causal direction represents complementary projection.

Complementary Projection and Learning Theory. A learning-theory approach to projection could operate on many levels of complexity. For example, it could be alleged that any child who hits upon the *tu quoque* accusation in dealing with parents ("But Mama, you and Daddy do it") gets less punishment than he would

TABLE 1

Relation between Reputation and Attribution as Predicted by Various Mechanisms

Reputation: A person seen by others as	Attribution: Tends to see others as			
	Complementarity	Similarity	Contrast	Reactivity
Anxious	Hostile, Critical	Anxious	Unanxious	
Bossy	Uncapable, Uninfluential	Bossy	Unbossy, Dependent	Unappreciative, Hostile, Stubborn
Critical	General unfavorability	Critical	Uncritical	Hostile, Touchy, Unappreciative
Dependent	Capable	Dependent	Undependent	
Hostile	Hostile	Hostile	Unhostile	Hostile
Obsequious	Influential	Obsequious	Unobsequious	
Secretive	Prying	Secretive	Unsecretive	Prying, Suspicious
Stingy	Greedy, Unappreciative	Stingy	Unstingy	Unappreciative, Hostile
Suspicious	Scheming, Unscrupulous	Suspicious	Unsuspicious, Gullible	Hostile
Touchy	Critical	Touchy	Untouchy	
Unscrupulous	Gullible	Unscrupulous	Scrupulous	Suspicious

otherwise. In addition, a child reared with siblings or in a neighborhood of children will be positively reinforced in many instances for saying, "It was not I, but he." Thus a generalized tendency to accuse others of one's own faults might be learned by reinforced trial-and-error behavior in childhood. This would result in the "similarity" projection of guiltworthy traits. However, the present emphasis is upon a simpler and more general aspect of learning corresponding to the social-perceptual theory already considered.

The learned traits of a person are the products of experience. Stimulus generalization and habit transfer predict that one will respond to newly met persons (or to photos of unknown persons) as though they were the kind of persons toward whom one's dominant response tendencies were originally learned. Thus, a person characterized by avoidance tendencies presumably learned these originally in contact with avoidworthy or punishing persons in childhood. He generalizes to novel persons by responding to them as though they too were avoidworthy or punitive. This exposition seems obvious for such responses as approach or avoidance. But the attribution of character traits to others involves a special kind of response which we can characterize as "judgment" or "depiction." To relate this to learning theory requires that, when considering articulate human beings, we explicitly expand our inventory of the relevant aspects of the original, learned response syndrome. Let us take a nonsocial case of one-trial learning. A child gets burned by a hot stove. This leads to an association of the stove with the responses of withdrawal and avoidance but, in addition and equally automatically, with the responses of "judging" or "perceiving" the stove to be "hot," "burning," "dangerous" or "bad." These "verbalizable descriptions" of the stove, these "attitudes toward" or "perceptions of" the stove, are an intrinsic part of what has been learned by the initial contiguity of the stove and pain. These features may not be overt in the initial event, but could usually be ascertained by the interview question, "Well, what did you learn from that?" The responses, "The stove is hot" or "The stove is bad," are perhaps even more likely than, "Stay away from the stove." These judgmental or perceptual verbal responses are just as apt to transfer to a novel stovelike object, as is the behavioral or locomotor avoidance response. When they do, they represent unwarranted projections.

As a more social example, consider a child to whom an adult has frequently caused pain. The original syndrome of responses learned to this adult includes not only avoidance and autonomic fear reactions but, in addition, the judgments or verbalizable perceptions that the adult is cruel, pain inflicting, bad, malicious, etc. If the child then generalizes the habits or response tendencies from this old stimulus to new ones, it will take the form of judging new adults to be cruel and malicious. The anxious timidity of the child will be projected as the perceived cruelty of adults. The complementary relation between the trait or response tendency and the perception or judgment thus exists in the original valid learning. It becomes an unwarranted complementary projection when this old learning is transferred to a new stimulus to which it is inappropriate. Thus the very pervasive learning processes of transfer and stimulus generalization predict an equally pervasive tendency toward complementary projection.

Similarity Projection

The most commonly assumed mechanism of projection is that of similarity projection: the trait projectively ascribed to the other is in reality one's own. The blanket assumption of similarity projection for all traits will be taken in this study as a major competing hypothesis for complementary projection. This will be done for reasons of simplicity and clarity without necessarily ascribing similarity projection, as used, to any one theoretical position.

The most common relationship in Freudian illustrations is one of similarity, accompanied by repression, or lack of insight into one's own traits. In many of the illustrations, however, traits are involved for which complementary and similarity mechanisms make the same specific prediction. In particular, this is true for the potent dimensions of love and hate. Thus, both mechanisms predict that those who perceive others as hateful will behave so as to be judged hateful. One has to consider traits like anxious, bossy, suspicious, and stingy, to generate differing predictions. This can be seen in more detail by comparing the similarity and complementary predictions presented in Table 1.

Love and hate, however, appear to be such basic dimensions of interpersonal interaction that the projective mechanism for all traits might understandably have been generalized from illustrations dealing primarily with these two traits. Since both projective mechanisms are potentially available as possible bases for such generalization, perhaps the wrong one has been chosen. Interestingly, many Freudian discussions do employ the complementary projection notion of misperceiving the world so as to justify one's own response tendencies. This may be particularly clear in the theory of transference in which the patient projects onto the therapist the attitudes of the patient's parent, rather than the attitudes of the patient himself, e.g., the guilt-ridden patient perceives the therapist as disapproving rather than as guilt ridden. The anxious patient perceives the therapist as hostile, not as anxious, etc.

The Freudian presentation may be ambiguous, however, when translated into the present research setting. One problem is its conceptual fragmentation of the person into multiple warring entities. Thus, if a person is guilt ridden (depressed, anxious) and perceives others as accusing and critical, this would be judged here to be an instance of complementary projection. In a Freudian presentation, however, it might be interpreted as the similarity projection of one fragment of one's self, i.e., one's hostile, self-judging superego. This does not seem essentially different from the learning-theory notion of generalizing from one's hostile, judging parents (who are internalized in the archaic superego) to contemporary adults.

Further ambiguity arises from the use of the term projection within Freudian theory. Whereas Freud (1955, p. 64) himself may have used the term projection more broadly, it has sometimes been used solely to denote a specific defense mechanism. In this latter usage, it has more consistently been given the limited similarity meaning accompanied with denial and lack of insight (Freud, 1937). Thus, while most of the responses elicited by a standard clinical tool such as the Thematic Apperception Test (TAT) may represent projections, a much smaller portion of these projections would represent the operation of the defense mechanism. Nevertheless, even in the case of the more restrictive defense mechanism usage, it could still be argued that the wrong mechanism has been chosen. For instance, in the classical psychoanalytic presentation of paranoia, latent homosexuality is characteristically interpreted as an underlying cause, and projection is the major defense mechanism used to reduce its threat. The paranoid represses his own homosexual tendencies and sees the world as full of homosexuals. The crucial question is how he sees these homosexuals as behaving. Does he see them as passive, latent homosexuals who like himself are awaiting or fearing seduction—similarity projection—or does he see them as actively trying to seduce him and thereby justifying his fears of homosexual assault—complementary projection?

Research putting the similarity and complementary hypotheses in competition is relatively lacking. Murray (1933) did not offer his subjects rating scales on the timidity or fearfulness of the photographed persons and thus gave similarity projection no opportunity to appear. Sears (1936) set out to confirm similarity projection and used an analysis which only allowed similarity, or its opposite, contrast projection, to appear. For his simple correlations between reputation scores and mean attributions, only contrast appeared (i.e., negative correlations) and this significantly so for the traits of orderliness and bashfulness. (His correlations for subgroups separated by degree of insight are interpreted as artifacts, as explained in the Analysis and Results section.) Hornberger (1960) put similarity and complementary projection into competition, using experimentally induced fear of electric shock. The subjects were given a forced choice among adjectives in describing photographs. Even though the fear induced was not associated with persons, complementary projection was found rather than similarity. (No projection, instead a suppression of both sets of projective adjectives, was found as a result of sex arousal.) Epstein and Smith (1957), measuring sex drive through reported sex activity, found support for complementary projection, although their setting was one in which similarity and complementary projection do not make clearly different predictions. Those men high in sex drive perceived photographed females as more sexually attractive to a highly significant degree. Unfortunately, they were not asked to rate the strength of the sex drive of the photographed persons (similarity projection). Probably this rating would have correlated very highly with the attractiveness rating. In a study which did put similarity and complementary projection into competition, Feshbach and Feshbach (1963) primarily found support for complementary projection but possibly for similarity projection as well. They aroused fear among boys and found the complementary projection of maliciousness onto photographs of men and boys. A weaker degree of maliciousness was also attributed to photographs of women and girls, but fright was not attributed to any of the photographs. In a second study maliciousness was only attributed to men, whereas the similarity projection of attributing fright to boys was almost significant. While the outcome most strongly supports complementary projection, there is the additional suggestion that the attribution of the complementary trait is selective. It is directed toward the type of target most capable of producing the subject's present motivational state. The frightened boys most consistently attributed maliciousness to men rather than uniformly attributing it to all targets.

Panglossian Projection

Some additional aspects of social perceptual theory such as the concepts of autism (Murphy, 1947) and perceptual defense (Bruner & Postman, 1948) can also be extended to the projective situation. These concepts agree with the pervasive common-sense notion of distorted perception known as "wishful thinking." It is here called *panglossian projection*, i.e., the tendency to perceive this as the best of all possible worlds. When applied to the relation between attribution and reputation, panglossian projection makes predictions which to some extent differ from those derived from either complementary or similarity projection. For negative traits it makes predictions opposite to complementary projection. Whereas complementary projection predicts that the anxious person will see others as forboding and hostile, panglossian projection predicts that the anxious person will see others as assuaging and friendly. For positive traits, panglossian predictions are less clearly opposite to complementary predictions. For instance, consider the unanxious, calm person. While complementary projection predicts that he sees others as helpful, friendly, and unhostile, the opposite relation of seeing others as threatening, hostile, and forboding does *not* follow from panglossian projection. Rather, the wishful, rosy view of the world should prevail. By arousing the negative trait of fright, Murray (1933) put panglossian projection into competition with complementary projection. Rather than seeing the unknown photographed persons as malicious or frightening, the girls could have projected the kind of traits that would make the world seem full of persons who would protect frightened children and who would assuage their need for reassurance. Thus, his outcome argues against panglossian projection.

On the other hand, if we demand a strict derivation from perceptual defense, it is not clear that panglossian projection should specify the particular content of the projection beyond insisting that it be pleasant. For instance, in considering the coed to whom the word PENIS is tachistocopically presented, perceptual defense only predicts her difficulty in "seeing" the word. It does not predict that her intrusion errors will be ones which specifically protect her from exposure to the threatening stimulus. Thus, the substitution of PANTS (which in this case functionally cover and hide the threatening object) is no more likely than other substitutions such as PENCIL, PENNY, PEANUT, or "I couldn't see anything." While these considerations allow greater leniency for finding support for panglossian projection, the Murray outcome nevertheless remains negative.

In considering reputation-attribution correlations, however, regardless of whether one is willing to assume specific threat-related wishful distortions or not, the representations of panglossian predictions are complicated by the lack of bipolar symmetry. As indicated, while it is plausible that anxious persons see others as nonhostile, it is not plausible on panglossian grounds that nonanxious persons see others as hostile. For these reasons, panglossian projection is not systematically represented in the present study.

Cassandran Projection

Another projective mechanism can be derived from social-perceptual theories of vigilance (Blum, 1954) in much the same way that panglossian projection is derivable from perceptual defense. When extended to the projective situation, vigilance corresponds to the misperceptions of the pessimist. Here, this tendency to see evil and misfortune everywhere is termed cassandran projection. However, the elaboration of cassandran predictions is made difficult for much the same reasons that made panglossian predictions unclear. On common-sense grounds, the projected evil should probably not be specifically related to the subject's own good traits. Rather, evil should be ubiquitously attributed regardless of which good traits the projector may possess. Furthermore, the lack of bipolar symmetry of predictions once again makes prediction of reputation-attribution correlations infeasible. Thus, cassandran projection, too, is not systematically tested in the present study.

OTHER MECHANISMS GENERATING PROJECTIONLIKE JUDGMENT ERRORS

Starting not from dispositional transfer theory, but looking at the possible sources of attribution-reputation correlations regardless of whether or not these be projective, two other processes generating such correlations must be noted. The hypothesized mechanisms are of such a nature that the attribution-reputation correlations generated would reflect contemporary stimulation rather than enduring misperceptions. They would reflect observations of the present peer group being judged which, relative to the judge, are valid and accurate. Projected content reflecting only the personality-forming residues of experience with past persons would be absent.

Contrast

For those traits referring to recurrent observable decisions, and for which bipolar descriptive terms are appropriate, application of an aspect adaptation-level theory (Helson, 1959) generates a prediction of contrast errors somewhat as it also predicts contrast illusions in psychophysical judgments. Take, for example, the continuum of thrifty-wasteful. In frequent major and minor decisions of purchase, discard, and use each person makes choices relevant to this continuum. The individual's typical threshold of decision on this dimension may be likened to a neutral point. Decisions deviating from this level in one direction would be regarded as too wasteful and in the other direction as too penurious; and decisions at the neutral level would be regarded as just right, neither penny-pinching nor wasteful. Expanding this to a group of persons observing each other's decisions, and validly noting the differences, the phenomenon of contrast projection is generated. Those at one extreme on the decision continuum will be epitomized by that polar description by their fellows. They in turn will characterize most of their fellows by the opposite pole. This phenomenon would seem most likely to occur for traits representing parallel or joint reactions of persons to a common environment, rather than reactions to each other as stimuli. It would seem likely to occur for such trait dimensions as orderliness, thriftiness, cautiousness-recklessness, persistence, industriousness, boldness-bashfulness, and the like. In addition, contrast errors would seem most likely to be found in the judgments of *familiar* persons. The judgments would represent valid relative knowledge of the position of these persons in relation to the perceiver, even though, as absolute judgments they would be invalid. This likelihood of stronger distortion in absolute judgments of well-known persons stands in contradiction to the previously presented projective mechanisms where greatest distortion is predicted for unknown persons or ambiguous situations. Thus, even in spite of Sears' (1936) prior designation of this type of error as a form of projection, we are somewhat hesitant to class it among the previously considered projective mechanisms. If, however, the contrast errors were indeed transferred from well-known to novel persons, contrast would belong among other varieties of projection utilizing dispositional transfer mechanisms.

In the most important prior study in this area, Sears (1936) discovered a tendency toward contrast projection for the traits of disorderliness and bashfulness, e.g., disorderly men tended to describe others as orderly. Murray (1938) found the same thing for ratings of the need for order by psychologists, this being the only strong projective correlation on the part of his raters which he obtained. Marks (1943) obtained a contrast correlation between judges' own skin color and the judgments made of others' skin color, providing essentially an adaptation-level or neutral-point interpretation.

Reactivity

A correlation between one's traits and the traits he sees in others can emerge from the reactions the others make to one's own traits. Thus, a domineering person may create reactions of avoidance on the part of his fellows. If he then judges his fellows to be more unfriendly than other judges do, this is not a false projection on his part, but a valid report of their reaction specifically to him. He reports his fellows differently than does the receptive person, because in reaction to his stimulus value they are in fact different. This mechanism is not technically projective, but our data-collection focus is upon the correlation between a person's reputed personality and the traits he ascribes to others. This is clearly a mechanism generating such a correlation.

For the traits of love and hate, the generated relationship produces a similarity correlation, and a "vicious-circle" mechanism, in which the reactions to the reactions are the same as the original response tendency—thus a hostile person generates hostility in others, to which his natural reaction is one of more hostility, in a positive-feedback relationship producing increasing amplitudes of reaction. For other traits existing in the person prior to experience in the present group, the reciprocal relationship may be more complex. The overly forward girl finds men timid, bashful, and undersexed, i.e., unattracted to her. The bossy person may be that way because of an initial tendency to view other persons as helpless, ineffectual, etc. (complementary projection), but by his bossiness generates in others responses he perceives as ungrateful and rebellious. The concept of reactive relationships among traits has high plausibility, but introduces the possibility of such complex reciprocal relationships that definite prediction becomes difficult. Adding further to the difficulty of prediction is the possibility that the same trait may be reacted to in totally different ways within different social groups. In one group people may be solicitous and understanding of the anxious person's plight. An anxious person in this group would rightly see others as unhostile and appreciative. Among another group, however, symptoms of anxiety may elicit ridicule, annoyance, and aversion. The anxious person in this group would judge others to be hostile, touchy, and critical.

In spite of these difficulties, some reactivity predictions are nevertheless presented in Table 1; and the theory is utilized in the discussion of results. Since it is hypothesized to reflect actual interaction with the persons being judged, it should show up more strongly in the judgments made of fellows than for judgments of photo persons.

INSTRUMENTS AND DATA COLLECTION

The list of 27 traits shown in Tables 2 and 3 was selected primarily to produce differential predictions for complementary and similarity projection. These predictions have been presented in Table 1. Thus, from the theory of complementary projection, anxious persons were expected to see others as hostile and critical rather than anxious; bossy persons to see others as incapable; dependent persons to see others as capable; stingy persons to see others as greedy; unscrupulous persons to see others as gullible; obsequious persons to see others as influential; secretive persons to see others as prying; suspicious persons to see others as scheming, etc. For hostile persons, the predictions of the complementary and similarity mechanisms are the same: They should see other persons as hostile. Contrast predictions were automatically represented as the opposite of similarity. In addition, the trait "orderly" was included as representing the strongest finding of past research. Reactivity predictions, also presented in Table 1, were represented only incidentally, and no trait terms were specifically selected for this purpose. The traits "masculine-feminine" and "sex-interest" were included for their relevance to the problem of identification. The rating on "good judge of personality" was included to support the general façade of the study, and in particular to justify the photo-judging task. Considerations of comprehensibility of trait terms and staff interests in particular traits were also involved.

In order to determine whether a population of college subjects would know the meanings of all the terms, a trait-definition test was administered to students in elementary psychology classes. The results revealed that several of the traits were of ambiguous meaning to the subjects. Traits such as "obsequious" and "insincere," "anxious" and "touchy," and "critical" and "capable" were frequently confused by the subjects. It was, therefore, felt necessary to define carefully all of the traits with particular stress placed upon the differential meanings of those traits between which definitional confusions occurred most often. Trait-definition sheets were constructed for this purpose, and these were included as part of the data-collection instrument battery. In addition, during the data-collecting sessions the examiners clarified the meaning of any traits that were still unclear to the subject after the trait-definition sheets had been read.

A 9-point scale was used for all ratings. A 9 was defined as "has this trait to a strong degree", 5 as "about average", and 1 as "almost totally lacks this trait." The subjects were encouraged to use as many scale points as possible in their ratings. To facilitate the use of a wide range of scale values the subjects were asked to rate in terms of the norms of their own living group, rather than in terms of the norms of the population as a whole. All ratings were done across traits, by individual or photograph. This procedure was used to maximize opportunities for projection to occur, since the alternative procedure of rating all individuals for each trait in turn would be more likely to result in equivalent attribution means for traits.

Complete sets of data were obtained from 236 women and 214 men, members of 19 residential groups. The groups were primarily Northwestern fraternities, sororities, and dormitories, but included living groups from a Catholic university and a teacher's college as well. Only those members of a residential group who were well acquainted with the other group members were allowed to participate. The subjects were paid for their time at a rate of $1.00 per hour, the typical payment being $5.00. These payments were either to the group as a whole, or to the individual, depending upon the group's decision.

The subjects were informed that immediately following each data-collection session all names on the rating sheets would be clipped off and replaced by code numbers, known only to the experimenter. This seemed to assure the subjects of the anonymity of their ratings, and no major problems in this respect were presented. While the overall level of cooperation of the subjects was high, the females did appear to be somewhat more conscientious in their ratings than the males. Groups 2–19 completed their tasks either in the living room or in some other large room of their living quarters, under the observation of the experimenter. The subjects were required to work independently and occasional reminders along these lines were necessary. The first group of subjects was permitted to complete the tasks in their rooms and the ratings of this group are probably not as independent as those of the other groups. In several of the dormitory groups some of the subjects stated that they did not know all of the participating group members well enough to give meaningful ratings. In all of these cases the subjects were encouraged to give ratings regardless of degree of familiarity. A few of the subjects claimed that they knew some of the other participating group members only by sight and in these cases ratings were not obtained.

There is no evidence that any of the subjects knew the true purpose of the experiment, and we may therefore assume that the façades used in the study (e.g., collection of "normative" personality data for college subjects, and ability to judge traits from photographs) were effective. Several of the subjects guessed that the first task involved the collection of reputational data for each individual, but none guessed (at least to our knowledge) that we were also interested in the attribution aspect of these ratings.

Three data-collection sessions were held for each group, with an interval of about a week between each session. In the first session each subject rated each of their residential fellows who were participating in the study, and in the same context rated themselves. Following this, they went over the ratings, indicating for all of the high ratings whether or not they thought the person was aware that he had this trait. The second data-collection session required ratings of 30 photographed persons of various ages and sexes. The photo booklets came from a previous research project (Campbell & Burwen, 1956) and supported the objective task assignment by containing uniform poses, two for each person, obviously made for this purpose rather than culled from news magazines or college annuals. In general, most of the subjects seemed to be fairly ego involved with this task, as witnessed by the frequent inquiries as to how accurately the photographs had been rated. The third session required a second self-rating, an estimate of the average rating one would receive from the other group members, ratings of one's family members, and various biographical and background information.

ANALYSIS AND RESULTS

Double-Standardized Scores

The first effort in data analysis was to compute, for each person, three scores on each trait: a Reputation score (his personality as rated by his peers), a Peer-Attribution score (his idiosyncrasy in the assignment of the trait to his peers), and a Photo-Attribution score. The computationally simplest approach would have been to use the mean of ratings received, and the mean of ratings attributed, for this purpose. Such simple scores have been used to check on major findings. It was felt, however, that such scores contained important sources of bias of predictable direction, and that they therefore did not represent the optimal test of the competing hypotheses. For this reason, for the basic analyses, we have employed double-standardized scores. Their rationale is as follows. In multiple-trait rating tasks there is usually found a very strong first factor or halo effect which leads all favorable traits to correlate positively with each other and negatively with unfavorable traits, etc. Thus it was possible that although we had 27 separate traits listed, we would be getting merely 27 repetitions of like-dislike. But for this general like-dislike dimension, our predictions from complementary and similarity projection are the same. Thus, if when a person is called "anxious," or "dependent," or "bossy," this means the same as calling him "hostile," no opportunity for complementary projection would emerge. For this reason, it was decided to convert each person's ratings of others for a given trait into a score which represented his general tendency to ascribe to others this trait to a greater extent than other unfavorable traits. This was achieved by the first standardization. All of a person's mean attribution values in the unfavorable direction were averaged (four favorably worded traits being reversed), and the trait-attribution means then converted into z scores representing deviations above and below this general tendency to ascribe unfavorable traits to others. Since at this stage individual differences in this *general* tendency were removed from the specific traits, this average unfavorability mean was itself entered as the twenty-eighth score. A similar transformation was performed for the reputation scores.

From previous studies and from general considerations it was expected that all projective correlations would be very low even where significant, and thus that it would take very large numbers of cases to establish the direction of the projective relationship. Since residential groups in which people know each other well enough to make trait ratings rarely have more than 35 members, it was obvious that data from many groups would have to be pooled. These groups could be expected to have average differences both in their true personality traits and in their vocabulary usages regarding personality traits. Our range of groups made likely some cultural differences, as there were included students at a Catholic university, girls in a private teacher's college specializing in the elementary grades, and members of both Jewish and Christian fraternities and sororities. If true differences between groups were reflected in accurate ratings by peers, this would result in a positive correlation between attribution and reputation on any trait on which the pooled groups differed. This would occur even if no projection were involved, i.e., even if all ratings were perfectly accurate. The resulting positive correlation between reputation and attribution would of course be erroneously interpreted as supporting similarity projection.

Such artifactual positive correlations can arise as follows. When all judges are judging the same persons, then the differences in their attributed judgments must be due to the judges themselves. But when different sets of judges are judging different persons, then attribution differences can be due to this difference in persons judged. Thus, even in the case of perfect accuracy, if several groups are pooled, with the judges and the persons judged coming from the same groups, group differences in attribution will be concomitant with group differences in reputations, resulting in positive correlations between attribution and reputation scores on a given trait. Analogously, if one regards the groups not as differing in real traits, but rather as differing in their willingness to ascribe certain trait terms, etc., the differences will appear in both attribution means and in reputation means, giving an irrelevant similarity correlation for the pooled population.

To avoid these irrelevant sources of an apparent similarity correlation, the scores of each person on each trait (both attribution and reputation) have been standardized in the more usual fashion around each residential group's own mean, and it is these double-standardized (*zz*) scores that have been employed in the pooled computations. This second standardization is regarded as achieving essentially the same result as if each projection correlation were computed separately for each residential group and these coefficients then averaged. While we believe these standarizations to be appropriate efforts to avoid real problems, it must be recognized that each standardization introduces problems of shared error variance among computed scores, through subtracting and dividing each through by a fallible observed (rather than independent and a priori), common value. However, while this affects the correlations *among* reputations and *among* attributions, it cannot have any artifactual effect upon the attribution-reputation correlations which are the main focus of inquiry. In all subsequent analyses the double-standardized scores have been employed, unless it is specifically stated otherwise.

Reliability, Validity, and Background Data

Table 2 shows the reliability coefficients for the major measures under consideration. These reliability measures are split-half values with the Spearman-Brown correction. For the reputation scores, the ratings by the odd code-numbered judges have been averaged and double standardized separately from the ratings by the even-numbered judges. Thus the expensive double-standardization process has been performed three times, once for each half and once for the pooled data. This seemed, however, the only appropriate way of estimating the reliability of the scoring process. In a parallel manner, the split-half reliabilities of the attribution scores have been computed by independent processings of ratings made by the subject of his odd- and even-numbered peers and of balanced halves of the 30 photographs.

In general, the reliabilities seem high enough to permit the observation of projection tendencies, even though both parts of the double-standardization process have made them lower than they would otherwise have been. The first standardization process has placed each subject's reputation (or his ratings of others) at an equal level of average unfavorableness, and thus removed a strong factor providing consistent individual differences. The second standardization has reduced the range of scores by making the means and variances of scores within all groups equal, a process further reducing computed reliability.

The reliability values are uniformly satisfactory for Peer Attribution. Among

TABLE 2

RELIABILITY, VALIDITY, AND BACKGROUND CORRELATIONS

Trait	Reputation reliability		Peer-Attribution reliability		Photo-Attribution reliability		Reputation × Self-Description		Peer Attribution × Photo Attribution		Self-Description × Peer Attribution	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
1. Anxious	.70	.74	.75	.84	.62	.67	.23	.23	.32	.43	.14	.28
2. Bossy	.83	.91	.65	.69	.43	.60	.47	.51	.06	.33	.09	.31
3. (Un)Capable	.85	.92	.88	.93	.81	.89	.34	.36	.19	.52	.18	.50
4. Complaining	.74	.80	.69	.74	.50	.58	.32	.34	.27	.23	.19	.05
5. Conceited	.83	.86	.70	.83	.66	.70	.40	.28	.31	.35	.02	.44
6. Critical	.75	.81	.70	.77	.50	.63	.20	.10	.19	.37	.14	.33
7. Dependent	.61	.79	.58	.81	.40	.66	.28	.34	.30	.40	.27	.27
8. Effeminate[a]	.89	.92	.77	.95	.65	.73	.40	.33	.30	.34	.54	.50
9. Greedy	.51	.64	.90	.83	.67	.74	.22	−.02	.22	.29	.28	.42
10. Gullible	.84	.88	.70	.81	.36	.48	.44	.31	.22	.23	−.05	.21
11. Hostile	.54	.65	.66	.84	.47	.76	.15	.26	.28	.42	.23	.40
12. (Un)Influential	.93	.94	.86	.92	.81	.88	.47	.46	.56	.57	.25	.42
13. Insincere	.47	.63	.74	.75	.69	.69	.21	.14	.31	.25	.22	.27
14. Obsequious	.56	.67	.76	.87	.63	.77	.27	.29	.43	.41	.27	.42
15. (Un)Orderly	.82	.89	.78	.85	.80	.91	.47	.46	.50	.56	.19	.16
16. Prying	.48	.60	.75	.84	.55	.75	.12	.27	.40	.24	.18	.44
17. Scheming	.67	.76	.77	.88	.58	.85	.23	.25	.34	.50	.31	.53
18. Secretive	.76	.78	.69	.79	.60	.74	.29	.22	.24	.34	.11	.30
19. Sex interest	.81	.87	.75	.90	.69	.77	.34	.23	.26	.29	.25	.52
20. Stingy	.51	.64	.68	.80	.65	.66	.34	.29	.26	.27	.03	.17
21. Strict	.83	.87	.66	.85	.66	.67	.39	.29	.25	.25	.30	.39
22. Stubborn	.55	.73	.78	.86	.62	.70	.18	.30	.38	.35	.16	.24
23. Suspicious	.40	.50	.72	.86	.69	.80	−.04	.18	.37	.39	.10	.34
24. Touchy	.47	.67	.74	.84	.56	.77	.16	.17	.37	.40	.28	.36
25. Unappreciative	.40	.47	.74	.78	.65	.77	.10	.04	.28	.42	.24	.30
26. Unscrupulous	.58	.61	.81	.88	.67	.86	.16	.02	.34	.40	.30	.52
27. (Un)Good judge of personality	.64	.80	.86	.93	.82	.90	.14	.12	.56	.61	.41	.62
28. General unfavorability	.84	.92	.93	.96	.86	.94	.14	.01	.53	.59	.60	.71

[a] For women this trait was labeled Masculine.

Peer Reputations the values show greater range and are undesirably low on crucial traits such as hostile. For these traits, it becomes important to consider "discriminant reliability," somewhat in the fashion of the multitrait-multimethod matrix (Campbell & Fiske, 1959). Is the reliability higher than correlations obtained between supposedly distinct traits? While these matrices are omitted from this report, inspection reveals that, in general, this is so. For hostile, the men's reliability is .54, and the highest correlation with another reputation trait is .49 with suspicious. For women, the reliability for hostile is .65, and the highest *r* is −.49 with gullible. Consideration of discriminant reliability for other traits shows that for men this criterion fails to be met for three traits: prying (*r* with scheming .49), unappreciative (*r* with unscrupulous .47, anxious −.43, and effeminate −.41), and good judge of personality (*r* with influential .65). For women, the trait insincere fails (*r* with conceited .64). All traits for both men and women meet this discriminant reliability criterion for both Peer Attribution and Photo Attribution, in spite of some low reliability values for male Photo Attribution.

The second half of Table 2 presents background data which will be of use in interpreting the measures and the results obtained with them. Some of these might be interpreted as validity data. For the reputation measure the only available independent assessment of the same traits are self-descriptions; and while these are far from ideal, it is of interest on several counts to note their correlation with Reputations. All values above .18 for men and .17 for women are significant beyond the .01 level (two-tailed test). By this criterion, three fourths of the correlations are significant. The correlations turn out to be of the same general order of magnitude as found for personality measures when independent assessments are checked against each other. For this literature (see Campbell & Fiske, 1959, for an extensive sampling) values of .50 are unusually high; and values of .30 are typical of "successful" measurement efforts. In a study of projection it would be particularly inappropriate to take self-descriptions at face value. Nevertheless, using .30 as reasonable reference value, it can be reported that all but 2 of the 20 correlations of .30 or greater meet the discriminant validity criterion of being higher than any of their heterotrait-heteromethod values (i.e., the correlations involving one of the measures with measures of other traits using the other method). While all heterotrait-heteromethod values have been computed, the complete matrices are not presented here. Unfortunately, among the traits for which the correlations between Reputation and self-description do not reach .30 is the important trait hostile. For men, the .15 "validity" value for hostile is exceeded by five heterotrait-heteromethod values, the highest (−.23) being between Reputation for hostile with self-description for obsequious. For women, self-described obsequiousness again provides the highest, and only, exception, −.27 compared with the validity of .26.

More legitimately interpreted as validity values are the correlations between Peer and Photo Attributions. It is of course crucial to the investigation that stable individual differences be found in projective tendencies, or idiosyncrasies in assigning trait terms to the average other person. The split-half reliabilities for attribution have, of course, answered this question affirmatively, but the demonstration of consistent idiosyncrasies, discriminantly specific to traits across such different target populations as live peers and photographs heterogeneous for age and sex, provides valuable support. Table 2 shows these validity values. For men, all but six of the traits meet the requirement of the validity values being higher than any heterotrait-heteromethod value. Bossy is obviously one of the exceptions, as are capable, critical, greedy, and secretive. For orderly, one trivial exception is found in that Peer Attribution of orderly correlates with Photo Attribution of influential −.51, to be compared with the validity value of .50. For women, for all but two of the traits the validity values are higher than all heterotrait-heteromethod values, these two being complaining and prying.

The final pair of columns in Table 2 shows the correlation between self-descriptions and Peer Attributions and are provided here only as background for the insight analyses to be reported below. Incidentally, they confirm for most traits the ubiquitous findings of a positive correlation. These highly significant values are often called assimilative projection (for a review see Lubetsky, 1960) and are not to be confused with the similarity projection under investigation in the present study.

Similarity Projection

Table 3 presents correlation coefficients which, if significantly positive, would provide confirmation of the hypothesis of similarity projection, i.e., the tendency of persons to attribute to others their own (as measured by reputation) traits. Of the 112 more or less independent values (men and women, each for double-standardized Photo-Attribution and Peer-Attribution scores) only 1 achieves significance in the similarity direction at the .01 level (two-tailed test). Bossy men perceive photo persons as bossy. Perhaps because not predicted, we are inclined to interpret this similarity relation as the 1 in 100 expected by chance in this many computations. All

TABLE 3

DIAGONAL PROJECTION CORRELATION VALUES
(CORRELATION OF REPUTATION AND ATTRIBUTION FOR SAME TRAIT)

Trait	Reputation (zz) × Peer Attribution (zz)			Average of group-by-group correlations: Reputation M × Peer-Attribution M			Reputation (zz) × Photo Attribution (zz)		
	Men	Women	Average	Men	Women	Average	Men	Women	Average
1. Anxious	.04	.03	.04	.06	.02	.04	.07	.03	.05
2. Bossy	−.04	−.03	−.04	−.03	.14	.06	.18*	.00	.09
3. Capable	−.01	.02	.00	−.05	−.10	−.08	.07	.03	.05
4. Complaining	.00	−.08	−.04	.02	.03	.02	.01	.03	.02
5. Conceited	−.09	.02	−.04	−.04	.06	.01	.06	.14	.10
6. Critical	−.11	−.05	−.08	−.06	.04	−.01	.06	.13	.10
7. Dependent	−.12	−.11	−.12*	−.06	−.11	−.08	.11	.02	.06
8. Effeminate[a]	.04	−.02	.01	.03	−.02	.00	.00	.02	.01
9. Greedy	−.08	−.04	−.06	−.12	.02	−.05	.09	.08	.08
10. Gullible	−.30**	−.21*	−.26**	−.20*	−.27**	−.24**	.13	.04	.08
11. Hostile	.04	−.03	.00	.02	.08	.05	.13	.00	.06
12. Influential	−.29**	.02	−.14*	−.24*	−.10	−.17*	−.08	.12	.02
13. Insincere	−.16	−.06	−.11	−.21*	.11	−.08	−.10	.09	.00
14. Obsequious	.02	.04	.03	.00	.01	.00	.08	.02	.05
15. Orderly	−.14	−.15	−.14*	−.12	−.26**	−.20**	.05	−.09	−.02
16. Prying	−.09	−.01	−.05	.00	.14	.07	.01	−.01	.00
17. Scheming	−.10	−.03	−.06	−.09	.10	.00	−.02	.08	.03
18. Secretive	−.17*	−.19*	−.18*	−.16	−.28**	−.22**	−.10	−.03	−.06
19. Sex interest	−.01	−.08	−.04	.00	−.07	−.03	−.04	.01	−.02
20. Stingy	−.11	.09	−.02	−.08	.00	−.04	−.02	.08	.03
21. Strict	.14	−.08	.03	−.01	−.14	−.06	.07	.11	.09
22. Stubborn	−.02	.05	.02	−.04	.11	.04	.06	.01	.04
23. Suspicious	−.25*	−.06	−.14*	−.11	−.02	−.06	.00	−.09	−.04
24. Touchy	−.10	.01	−.04	−.05	.06	.00	−.03	.13	.05
25. Unappreciative	.05	−.17	−.06	−.06	−.06	−.06	.03	−.15	−.04
26. Unscrupulous	.00	.01	.00	−.11	.08	−.02	.07	.01	.04
27. Good judge of personality	−.15	.09	−.03	−.03	−.04	−.04	−.10	.10	.00
28. General unfavorability	−.03	.03	.00	.00	.06	.03	.04	−.11	−.04

[a] For women this trait was labeled Masculine.
* $p < .01$ level (2-tailed test, single value).
** $p < .000063775$ level (2-tailed test), corresponding to the full-matrix $p < .05$ level, as discussed in the text.

of the other significant values are negative, confirming contrast projection. These significant contrast values all come from Peer Attribution and tend to be parallel for men and women.

The general absence of similarity projection goes against the common assumption of current psychology (particularly when reconfirmed with insight taken into consideration as reported below). On most traits, this absence is congruent with the theoretical expectations which lead to this study in that complementary projection was seen as more directly following from social cognition and learning theories. However, for the traits of general unfavorability and hostility, complementary projection and reactivity also predict a positive value; and the negative results on these traits are most disappointing.

These results for hostility and general unfavorability are so out of keeping with our own and others' expectations that it was judged desirable that they should be checked independently of the double-standardization process. Particularly for hostile, the first standardization might have removed too much, as it was done to prevent other traits from being repetitions of the unfavorability component which is validly a part of hostile. For this reason, the similarity-contrast Peer-Attribution correlations have been computed separately for each of the 19 groups, using simple mean ratings for both reputation and attribution rather than double-standardized values. For each trait, the averages of these correlations are shown in the second set of columns in Table 3. While differing in detail, the results are judged to confirm the double-standardized analysis.

Since the expectations for hostile were so strong, it seems well to inspect the group-by-group values for this trait in more detail. As expected due to artifact, and as shown in Table 3, this analysis moved the results in the direction of a similarity outcome for the women's groups, even though the mean value remains unconvincing. For 7 of the 10 women's groups the values were positive (the largest being .25), for 2 the values were negative (largest −.22), and for 1 the value was 0. By a one-tailed sign test, the 7 confirmations out of 9 give a $p < .09$. Male groups split 4:4 with one 0 (range from .44 to −.32). The combined sign test gives 11 out of 17 ($p < .17$). These outcomes may represent some meager encouragement of the similarity hypothesis; although, as noted, the same prediction is also made by both complementary projection and reactivity. Such sign tests similarly provide slight support for the general unfavorability means. For women, the score is 8+, 2− (range from .19 to −.30; $p < .06$) for men, it is 5+, 4− (range from .35 to −.48). The combined 13 confirmations out of 19 give $p < .09$ (one-tailed test). Considering the very large numbers of cases involved in the study, these slight trends probably should not be taken as confirmatory, especially when the exceptions are of such large magnitude and the average values resulting are so small. In searching for an explanation of these disappointing findings, it is tempting to point back to the low reliability values and questionable validity for hostile, particularly for men. This seems unwarranted, however, in that the results for general unfavorability are only slightly less disappointing, in spite of the fact that the reliabilities are substantially better.

Insight Analyses

Similarity projection, as we have represented it so far, may be a useful reference point, but is not a theory actually advocated by anyone. To represent Freudian or paranoid projection, it should be qualified by considerations of insight. According to this position, it is hostile persons lacking insight into their own hostility who should project hostility onto others. Perhaps the prediction would hold for such persons even if not for the whole population. In the present study, after rejecting prior approaches, we have operationalized insight in three ways.

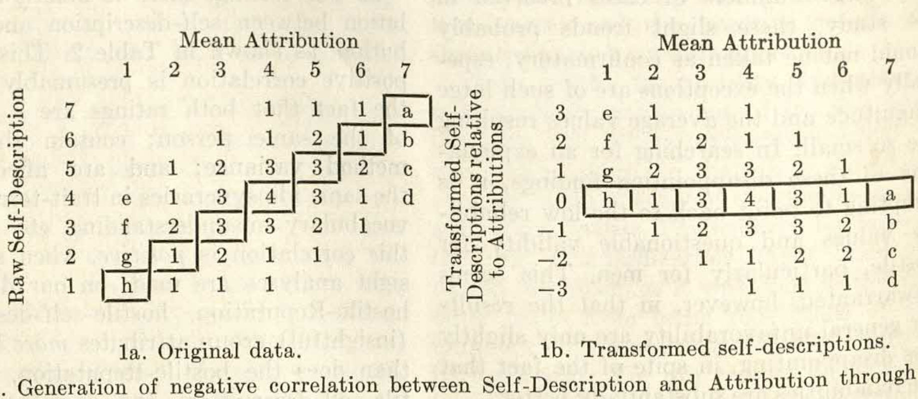
Rejected Approaches. One approach represented in the previous literature is to use the agreement between self-description and reputation as a measure of insight. Some seemingly direct procedures in this regard are, however, highly misleading. Let us take two groups rated hostile, one of which describes themselves as hostile, and the other describing themselves as nonhostile. The second group would be expected to attribute more hostility to others, since it is hostile (using reputation in lieu of psychiatrist's ratings), but lacks awareness of this fact. Note, however, that these two groups not only differ on insight, but they also differ, more simply, on self-description. As Wylie (1961, pp. 295–298) points out, if self-description were in a direct way correlated with Attributions, this correlation would produce Attribution differences without any need to introduce considerations of true personality, reputation, and insight.

In our setting, there is indeed a correlation between self-description and Attribution, as shown in Table 2. This clearly positive correlation is presumably due to the fact that both ratings are a product of the same person; contain the same method variance; and are affected by the same idiosyncrasies in trait-term usage, vocabulary misunderstanding, etc. Because this correlation is *positive*, when such insight analyses are made on our data, the hostile-Reputation, hostile–self-description (insightful) group attributes *more* hostility than does the hostile-Reputation, nonhostile–self-description (no insight) group.

This complete reversal of similarity projection theory we interpret as solely a function of the simple correlation between the self-descriptions and the Attributions, with Reputations (and hence insight) remaining irrelevant to Attributions as indeed the whole-group correlations of Table 3 show them to be.

Building upon previous criticisms by Rokeach (1945), Zimmer (1952), and Wylie (1961), we interpret the "confirmation" of similarity projection by the noninsightful, previously reported by Sears (1936) and also in part by Rokeach (1945), as purely a function of the direct self-description and Attribution relationship. Their confirmation is totally independent of Reputation and insight. In neither study do they report the direct correlation between self-description and Attribution, but due to their method of transforming the self-descriptions before using them, a *negative* correlation would inevitably be produced. This artifactually produced negative correlation automatically insures finding similarity projection among those lacking insight. *Noninsightful* persons are defined as those high in Reputation but low in self-description. If self-description is negatively correlated with Attribution, those lacking insight will tend to attribute the trait to others. Likewise, those possessing insight (high in self-description) will tend to attribute little of the trait to others.

Because of the complexities involved, it seems desirable to spell out in some detail the way in which their transformations produce negative correlations between Attribution and self-description. Let us suppose that in terms of raw self-ratings and mean Attributions a positive correlation existed, as is found in most instances. Figure 1a shows an artificial scatter plot for such a case. The numbers represent frequencies. The letters illustrate the location of some specific persons in the two scatter plots. Both Sears and Rokeach translate the self-ratings into self-ratings relative to Attributions, calling a self-rating high only if it is high relative to the attributions the person gives. Attributions, on the other hand, are used as a final dependent variable in their raw form. We have tried to illustrate such a process in Figure 1, where the identical-rating diagonal of Figure 1a defines a horizontal divider for the transformed self-descriptions of Figure 1b, defining a transformed value of zero on self-descriptions. Thus, for example, the self-description of Person e is converted from his raw score of 4 into a transformed score of 3, since his self-description exceeds his mean attribution of 1 by that amount. His data point is then entered in its new proper location in Figure 1b. Likewise, Person d, also with a raw self-description of 4 but a mean attribution of 7 receives a transformed self-description value of −3. His transformed data point is located in the lower right corner of Figure 1b. When all the data points are thus transformed and replotted, the resulting scatter plot (depicted in Figure 1b) clearly shows a high negative correlation between transformed self-descriptions and mean Attributions. Or to put it another



1a. Original data. 1b. Transformed self-descriptions.

FIG. 1. Generation of negative correlation between Self-Description and Attribution through scoring Self-Descriptions relative to Attributions.

way, we are correlating Attributions with self-description minus Attribution, thereby generating an index correlation which must be negative under all conditions in which self-descriptions and Attributions are not perfectly correlated.

With such a negative correlation assured, Sears' results now become predicable through attention to self-descriptions only. Reputation and insight become needless considerations. Those who describe themselves (relative to their Attributions) as not obstinate tend to see others as obstinate. This happens in Sears' data whether or not they are obstinate in Reputation. But when they are obstinate by Reputation they are classified as noninsightful similarity projectors, and when they are nonobstinate by Reputation they are classified as insightful contrast formationists. This two-process interpretation is here rejected as unparsimonious.

Rokeach (1945) makes essentially this criticism of Sears' method in one packed sentence:

> It might be expected purely on mathematical grounds that those who rate themselves lower than they rate others must necessarily attribute a higher ... score on the whole to others than those who rate themselves higher than they rate others [p. 160].

Through a complex procedure, he eliminates much of the error in his major analysis of his own data and hence finds no evidence of similarity projection. But in the one analysis providing more seemingly positive results, Rokeach (1945) uses self-description relative to Attribution and obtains conclusions interpretable as due to this alone, with reputation and insight needlessly brought in, e.g.,

> Beautiful women who ... underestimate their own beauty attribute more beauty to others than do beautiful women who ... overestimate their own beauty [p. 164].

Wylie (1961) provides a more detailed treatment of Rokeach's analyses.

Use of Self-Description to Determine Insight. These two forms of a misleading use of self-descriptions to assess insight have thus been rejected. Where self-description and Attribution are correlated directly, any comparisons of the Attributions of insight groups which differ on self-description must be ruled out. But if we start with insight groups that are equated on self-description, but differ on Reputation, such difficulties are avoided, at least for those variables which have a zero overall correlation between Reputation and Attribution. Wylie (1957, 1961, pp. 295–298) has employed such analyses. Thus, of two groups describing themselves as nonhostile, those persons hostile in Reputation can be said to be both hostile and lacking in insight and should therefore project more hostility onto others than those both self-described and reputedly nonhostile (i.e., both insightful and lacking in hostility). This prediction would seem on a priori grounds to be an even more nearly optimal representation of the denial-projection mechanism than the comparison between two reputedly hostile groups. This prediction has been tested by taking from the low third on self-described hostile the top and bottom third on Reputation for hostile. These two groups were then compared in their Attributions of hostile. No hint of significant differences was obtained. For Peer-Attribution the t ratios were men, $+.310$ ($df = 42$; plus sign indicating correct direction) and women, $+.522$ ($df = 42$). For Photo-Attributions: men, $t = +.082$ ($df = 42$) and for women $t = -.801$ ($df = 42$). This analysis, comparing two groups describing themselves favorably, one lacking and one having insight, was done for each trait, for both types of attributions, and for each sex. In only one of these 112 t's was the $p < .05$ value reached for a difference in the similarity projection direction. This was for women, Photo-Attribution, touchy, $t = 2.487$ ($df = 42$). The other eight values exceeding $p < .05$ were in the contrast direction, primarily reflecting the contrast correlations of Table 2 and will be discussed below. Due to a lack of consistent pattern even for touchy, the total absence of the process on equally plausible variables such as hostile, stingy, stubborn, etc., and the occurrence of t values in the similarity direction at less than chance expectancy, these results cannot be regarded as confirming the similarity-projection-with-lack of-insight hypothesis.

While persons who described themselves at the unfavorable extreme scarcely fit the denial-projection paradigm, a parallel analysis starting from this extreme has likewise been made. For this, none of the 112 values produced a t at the $p < .05$ level in the similarity projection direction, while five reached that level in the contrast direction. This essentially equivalent outcome for self-description's extreme in either direction supports a contrast relation as the strongest finding.

Use of Estimates of Own Reputation to Determine Insight. On the final data collection task, the subjects were asked to estimate the average rating which they would receive from their peers. These guessed-other ratings are in general similar to the self-descriptions, but deviate from the self-descriptions in being more in agreement with Reputations, significantly so in many instances. A person who regards himself as nonhostile, but who is aware of his reputation for being hostile, might not be regarded as wholly lacking in insight.

The guessed-other ratings have been employed in a manner analogous to the use of self-descriptions. Two groups have been compared, both guessing that others describe them favorably, one with favorable Reputations (insightful), one with unfavorable Reputations (noninsightful). Of the 112 such comparisons, only one reaches the $p < .05$ level in the direction favorable to the noninsightful similarity-projection direction. This was for women, Peer Attribution, general unfavorability, $t = 2.283$ ($df = 42$). The trend for men was almost significant in the opposite direction. Six comparisons reached significance in the contrast direction.

Again, although it makes little sense to regard as lacking in insight those who describe themselves unfavorably, a parallel analysis was done comparing two groups expecting others to describe them unfavorably, one group correct in this, the other incorrect (noninsightful). Of the 112 such comparisons, only one was in the similarity-projection-where-insight-lacking direction. This was for women, Photo Attribution, scheming, $t = 2.323$ ($df = 31$). In this analysis, 11 comparisons reached significance in the contrast direction. With fewer similarity-with-noninsightful comparisons significant than would be expected by chance, and with these inconsistent in pattern, the outcome is interpreted as rejecting the hypothesis. Wylie (1957) in a somewhat similar analysis likewise found no evidence of projection. Once again, if anything, these results support contrast.

Reputation for Insight. At the end of the first task, peer rating, the subjects were instructed as follows:

> Some people, although they possess a given trait, are not aware that they do. Now that you have finished making your ratings on the various traits, go back to those ratings that are 6 or higher. In all boxes which contain ratings of 6 through 9, indicate whether the person is aware or unaware that he possesses the particular trait. If *aware*, put a small *a* in the box. If *unaware*, put a small *u* in the box. If you don't know whether or not the person is aware that he possesses the trait, leave that box blank. Be sure to consider every box with a rating of 6 or higher.

The task was included as a part of our effort to obtain all feasible measures of insight compatible with group data collection. It was felt that for both clinician and casual observer, the judgment as to presence or absence of insight often accompanies the initial observation of the presence of the trait (e.g., "She doesn't realize she's being selfish."), rather than being a comparison process combining separate observations as to true character and self-image. Furthermore, as Peer Reputation has in this study substituted for psychiatrist's judgments as to true personality, Peer Reputation for having insight is being substituted for the psychiatrist's judgment as to the presence or absence of insight.

On semantic grounds, it was regarded as too confusing to ask for insight ratings where the person lacked the trait, i.e., was given low ratings. Since most of the traits were unfavorable, the dynamically relevant pole was covered for them. The rating was made a separate task to avoid confusing the basic ratings. While the instructions gave those who were lazy an opportunity to evade the task, it was felt better to base the classification upon judgments made by confident and conscientious raters than to force judges into decisions which they did not understand. The comparison groups were of necessity high in their reputations on the traits in question. From among these, those receiving six or fewer mentions of either sort were eliminated. Each of the remaining persons received an insight score based upon the percentage of mentions indicating insight. For each trait, the upper and lower thirds on reputation for insight were compared as to their Peer Attributions and Photo Attributions. Due to too few cases, the comparison could not be made for all traits. In the 104 comparisons made, in only 2 did the value reach the $p < .05$ level, and even then, just barely. One of these cases was in the similarity-projection-with-lack-of-insight direction. This was for men, Peer Attribution, secretive, $t = 2.095$ ($df = 28$, $p < .05$). There is no consistent pattern, the difference for women being in the opposite direction. Again, the occurrence of significant values is less than chance expectancy. Again we interpret the results as completely discouraging for a similarity projection hypothesis even when qualified by considerations of insight.

Contrast Relationships

As indicated in the introductory discussion, a contrast relationship for the trait orderly is the best established relationship to be found in the previous studies. We judge this finding to be confirmed in the present study. For men and women combined the .14 value exceeds the .12 required for the $p < .01$ level (two-tailed test) and for each separately ($p < .05$). The procedure of computing the correlations with unprocessed scores from simple means enhances the magnitude of the effects for women and for the combination of men and women. It seems an appropriate approximation to use for these average r's a significance level computed in terms of the total N of 450, in which case the value of .20 becomes significant at .000063775 (a level chosen for reasons to be discussed in the next section).

These outcomes, combined with the earlier findings of Sears (1936) and Murray (1938), seem to establish firmly a contrast relationship for orderly. We judge this to be so. And yet the finding is not so strong as to be counted on in every social group of 20–30 members. The correlation was in the contrast direction for only 5 of the 9 men's groups (range: −.59 to .24), and 9 of the 10 women's groups (sign test $p < .02$, one-tailed; range: from −.53 to .29). For the combined score of 14 out of 19, $p < .04$ (one-tailed test).

As can be seen from Table 3, all of the other significant relationships for Reputation × Peer Attribution are likewise in the contrast direction. For gullible, the relationship seems firmly established. Even here, however, not all individual groups show the effect. For men, the score is 6 contrast, 2 similarity, 1 zero (range: from −.54 to .23). For women, 9 contrast, 1 similarity (range: from −.59 to .23). For the combined 15/18 confirmations, $p < .01$ (two-tailed test since direction not predicted). Contrast also seems well established for secretive, which has the best score of all by the sign tests, contrast showing in all 10 of the women's groups (range: from −.10 to −.48) and 7 of the 9 men's groups (range: from −.46 to .04). For this 17/19 score, $p < .001$, by a two-tailed test. Contrast may be present also for dependent (probably as a pale replicate of gullible), influential, and suspicious.

Note that there is no consistent tendency for these contrast relationships to be replicated in the Photo-Attribution data (e.g., for gullible, the slight trends are in the opposite direction, for secretive they are in the same direction, and for the others they are mixed). The reliability and validity indicators of the Photo-Attribution scores do not explain this, except possibly for gullible, where the Photo-Attribution reliabilities are low, as shown in Table 1. This absence of contrast in the Photo Attributions is quite consistent with the a priori theoretical expectations. As explained in the introduction, the dispositional theory which seems most clearly to generate contrast relationships is adaptation-level theory. In terms of this theory, the judgments reflect valid knowledge of the differences between the judge and the person judged. Such differences are reported, however, in terms of a judgmental language in which each judge places himself at the neutral point of his scale. On these grounds, one would most specifically expect contrast relationships in judgments of persons whose behavior the judge had had the opportunity of observing. If this interpretation is correct, the contrast relationship perhaps should not technically be called projection, since it is not purely a product of the baggage of dispositions which the judge himself brings into the present situation. Rather, it also reflects valid recognition of the specific self-other differences actually a part of the present situation.

As has been mentioned, the *t* tests of the insight analyses repeatedly turned up contrast results. Most of these reflect the relationships already shown; and where they do not, they lack a magnitude or consistency of pattern sufficient to demand interpretations of exposition. For women on dependent, contrast in Peer Attribution occurs most strongly when self-descriptions are high. The same trend exists for men and is significant ($p < .05$) for both men and women when guessed other is high. On insincere, for men, contrast occurs for both Peer and Photo Attributions when self-description is high. On unappreciative, for women, contrast occurs for both Peer and Photo Attributions when self-description is low.

The trend for Photo Attribution to parallel Peer Attribution is greater in these *t* tests comparing extremes than in the overall correlations of Table 3. To the extent that contrast errors are found in the Photo Attributions, and therefore represent a stable disposition brought to the judgment situation, it makes sense to consider contrast as a legitimate form of projection. On logical grounds alone, there is no reason why contrast could not be dispositional. Indeed, these analyses suggest that this may sometimes be the case for persons who describe themselves as extreme. This makes sense in that such persons are most likely to have noticed or learned that a consistent unidirectional difference exists between themselves and others in a variety of situations. Thus, while we would argue that nondispositional contrast errors would more frequently appear among those who describe themselves as average or middling, dispositional contrast would be more likely in those who see themselves as extreme.

In addition to the instances just cited, Photo Attributions join Peer Attributions in showing contrast for men on influential (when either self-descriptions or guessed other lie in the middle third) and on orderly and insincere (when guessed other in middle, similar trend for middle self-descriptions). While if replicated and studied in their own right some of these complications might prove both stable and meaningful, they are not judged to add interpretably at present to the overall correlations of Table 3.

The adaptation-level theory of contrast judgments has implications not only for the simple relationship between reputation and attribution, but also for the qualification of this relationship by self-description. We have argued that the contrast relationship occurs because persons use their own behavior as a neutral reference point in reporting on the actual differences between their behavior and that of others. As noted, for persons who most frequently exemplify this process, we might expect self-descriptions to lie in the middle range, rather than being extreme. For this reason, we have accompanied the self-description insight analyses reported above with a parallel set of *t* tests which compare two groups both lying in the middle third on self-description, one group high, the other low on Reputation. Since equal sized groups are involved, one can compare the magnitudes of these *t*'s for high and low self-descriptions computed for the insight analyses reported above. The middle self-description *t*'s should be largest. For men, Peer Attributions, this is so for 15 of the 28 traits, where by chance 9.3 would be expected ($p < .02$ by a one-tailed sign test). For women, the score is 12 out of 28 ($p < .20$). For the specific traits showing contrast most strongly, the results are as follows: for gullible there is confirmation for men (low 1.60, middle 3.30, high 2.75, all $df = 42$), but not for women (2.62, 2.38, 1.61, all $df = 42$). For men, influential, the pattern is strong (2.25, 4.26, 2.09), as it is also for men suspicious (1.46, 2.08, 1.30). For orderly, the trend is present for men (−.71, 1.84, 1.24), but not for women (2.73, 1.38, 1.24). For secretive, the trend is exactly reversed for both men and women with the middle group showing least contrast (men: 1.99, −.01, 1.00; women: 2.59, 1.01, 1.30). (None of the differences between t's reaches a $p < .05$ level of significance, which would require a difference between t's of about 2.82). The trends for men thus offer some support of the expectation, although far from definitive. In research organized solely around adaptation-level theory, it would be better to use bipolar rating scales (e.g., hot-cold, heavy-light, sloppy-orderly, etc.) rather than single-term trait designations, so that a phenomenal neutral point could be more clearly designated. Intermediateness or averageness such as here employed is not optimal for this aspect of the theory.

Complementary Projection and Other Off-Diagonal Relationships

The "projection matrices" of this project consist of 28 Reputation variables as column heads and 28 Attribution variables as rows. The diagonal of these matrices represents the tendency of persons exceptional on a given trait to attribute this same trait to others. These diagonal values are those presented in Table 3. Except for the few that overlap with similarity projection, the predicted relations for complementary projection and reactivity lie off this diagonal in that the projections are expected to have a trait content that differs from the judge's dominant personality characteristic.

Predicted Complementary Projection. The trait selection was predominantly directed so as to put complementary projection into competition with similarity projection. As it turns out, the outcome for the specifically anticipated complementary projections is almost as disconfirming as the outcome for similarity. Here are the outcomes for the less abstruse predictions: anxious persons were expected to see others as critical (+, +.05, −.04, +.02, +.05) and hostile (+, −.01, +.02, +.07, +.09). (The shorthand in the parentheses gives first the sign which would be regarded as confirmatory, followed by the four correlations obtained, in the order Reputation × Peer Attribution for men and women, Reputation × Photo Attribution for men and women.) For significance at the .05 level (one-tailed test) a value of .11 is required, though due to the large number of cases used and the large number of hypotheses tested this value seems relatively meaningless. Values of .18 and .17 are required for significance at the .01 level (two-tailed test). Bossy persons were expected to see others as not capable (−, −.15, +.04, +.06, −.01). This expectation was based upon the mother-in-law model, in which a misperception of others as incompetent leads to excessive advice giving. In the nonhierarchical peer situation, our ratings perhaps described a more exploitative bossiness, based upon the misperception of others as gullible (but see below). Touchy persons were expected to see others as critical (+, −.05, −.07, +.04, −.01); dependent persons to perceive others as capable (+, −.05, +.01, −.10, +.04); stingy persons to see others as greedy (+, −.06, −.01, +.04, +.04); obsequious persons to see others as influential (+, +.06, +.06, +.07, +.10); secretive persons to see others as prying (+, .06, −.18, .00, −.03); suspicious persons to see others as scheming (+, −.11, −.01, +.12, −.11) and unscrupulous (+, −.04, +.11, −.05, +.03); and critical persons to see others with general unfavorability (+, −.02, +.19, +.04, +.03). While the best of these values are better than the best similarity confirmations, the lack of consistent pattern and the low absolute magnitudes lead us to interpret them as nonconfirmations.

Significance Levels for the Best of Multiple Tests. The complementary predictions were made on common-sense considerations as to the kind of behavior which might be concomitant with given misperceptions of others. However, the state of personality or dispositional theory is hardly such as to provide rigorous deductions of such relationships. Nor are these by any means the only plausible off-diagonal relationships. There are thus grounds for looking at the high correlations and then judging which explanatory mechanism is most plausible for each.

With 784 correlation coefficients in each of four matrices, the opportunity for overinterpreting chance variations of a true value of zero makes inappropriate the use of the standard criteria we have been using so far, where the loci of the relationships to be inspected have been specified in advance. While we know of no entirely satisfactory solution to the problem, several base lines for comparison have been examined and are presented in Table 4. First the number of correlations "significant" by the usual single-value criteria are examined. These tests form the first section of Table 4. The first row indicates the correlation magnitude required for significance at the .05 level for men and women. Entered in the second row are the number of correlations which exceed the .05 level for each matrix. Among 784 independent samplings from a true correlation of zero, one would expect 39.2 values to exceed the $p < .05$. The chi-square test of obtained against expected with $df = 1$ was computed for each matrix and appears in the third row. Results of an analogous procedure for the number of correlations exceeding the .01 level are also presented. Note that the chi-square values are significant well beyond the .01 level for all matrices except men's Photo Attributions where they fall far below.

The assumption underlying this type of test is, of course, inappropriate. The 784 correlations are not independent, any given array of scores being reused 28 times, and the arrays for separate variables frequently correlating highly with each other. It could be argued that this lack of independence underestimates significance through providing in actuality less than 784 opportunities for a chance deviation to exceed a given value. The expected values might thus be uniformly too high. On the other hand, a sampling error generating a high correlation in the obtained data might be expected through the correlated sampling error to be represented in several obtained values, thus inflating the obtained count and overestimating the significance level. Were the true values to lie away from zero, however, such correlated sampling error could lead in a parallel way to underestimation. Block (1960) has demonstrated for one setting a dramatic underestimation of significance in a multiple-tests problem. His problem was analogous to the evaluation of significance levels for but one column of a table such as ours, and his solution would be analogous to generating a null comparison through correlating the obtained row arrays with random number values for the column variable. This solution does not seem applicable in detail to our setting of nonindependence both among the arrays constituting column heads and among those constituting row variables. Generalization of his outcome, however, seems in part justified and would imply that the tests of Table 4 are unduly pessimistic.

Another approach has been to look at the variance expected for a true correlation of zero, repeatedly sampled with the N of 214, or 236. (See the second section of Table 4.) The departure from expectancy has been tested for significance with a chi-square and a t test (Walker & Lev, 1953, p. 184). Here again, three of the matrices significantly exceed the expected variance, while men's Photo Attribution does not. For this test, the lack of independence would not seem to have any effect upon the expected value, but could suppress or enhance the obtained variance and thus make the sampling fluctuation of σ_r larger than that assumed by the formula. But insofar as the test is relevant

TABLE 4

TESTS OF WHOLE PROJECTION MATRICES FOR SIGNIFICANT CORRELATION

	Reputation × Peer Attribution		Reputation × Photo Attribution		Men + Women (N = 450)
	Men (N = 214)	Women (N = 236)	Men (N = 214)	Women (N = 236)	
Tests of number (#) of "significant" correlations					
r for p = .05 (2-tailed test)	.134	.128	.134	.128	.092
Obtained # > .05 (expected = 39.2)	101	92	51	72	
χ^2 for # > .05	102.6*	74.9*	3.7*	28.9*	
r for p = .01 (2-tailed test)	.177	.168	.177	.168	.122
Obtained # > .01 (expected = 7.84)	19	29	7	18	
χ^2 for # > .01	16.2*	57.7*	.4	13.3*	
Tests of variance of correlation matrix					
$M\ r$	.0008	.0007	−.0008	−.0015	
σ_r for $r = 0$	.0685	.0652	.0685	.0652	
Obtained σ_r	.0829	.0812	.0690	.0748	
z	8.28*	9.71*	.29	5.91*	
Tests for single correlation values					
r for p = .000063775 (p = .05)	.275	.262	.275	.262	.189
r for p = .000012765 (p = .01)	.299	.285	.299	.285	.206

* $p < .01$.

it again reassures us that some of the high values in at least three of the matrices are interpretable as nonchance events. Which ones are to be so interpreted remains in doubt.

Because of the generally low values obtained for attribution correlations and because of the questionable assumptions required by our significance tests for whole projection matrices, one might still question whether there is any relation between attributions and reputations. In view of the large number of correlations in each matrix, perhaps even the previously discussed contrast correlations represent chance deviations around an average r of zero. Canonical correlation provides a test of whether two sets of variables are significantly related while making no assumption of independence between the components of each set (Cooley & Lohnes, 1962). It is analogous to a multiple correlation in which there are multiple criteria as well as predictors. It determines the ways in which components of the two sets can be combined and weighted so as to maximize the correlation. Canonical correlations were computed between double-standardized Reputations and Peer Attributions for both females and males.[5] For females, the maximum canonical correlation was .73 ($p < .0001$). In addition to this first relation between reputations and attributions, there are two other significant combinations among the two domains. In deriving the successive canonical variates, each new combination of weights within a domain is orthogonal to other previous combinations. The second highest canonical correlation of .67 is significant at $p < .0004$; the third is .64 ($p < .03$). For

---

[5] These analyses were supported by a Research Grant from the United States Office of Education Project C-998, Contract 3-20-001, under provisions of Title VII of the National Defense Education Act.

TABLE 5

TRAITS CONTRIBUTING THE LARGEST WEIGHTS TO THE RELATED REPUTATION-ATTRIBUTION CANONICAL VARIATES

Canonical correlations	Reputation weights		Attribution weights	
Females				
No roots removed	Uncritical	.43	Influential	.49
$R_c = .73$	Orderly	.39	Unorderly	.41
$\Lambda = .0047$	Scheming	.36	Dependent	.39
$\chi^2 = 943.7$	Bad judge of personality	.26	Unscheming	.32
$df = 729$				
$p < .0001$				
One root removed	Unprying	.37	Influential	.36
$R_c = .67$	Unstubborn	.37	Ungreedy	.33
$\Lambda = .0102$	Strict	.35	Appreciative	.32
$\chi^2 = 809.2$	Hostile	.33	Orderly	.30
$df = 676$	Bossy	.33	Unstrict	.28
$p < .0004$	Suspicious	.26	Unscheming	.23
	Unorderly	.25	Unsecretive	.23
	Bad judge of personality	.25	Feminine	.23
Two roots removed	Unscrupulous	.51	Orderly	.48
			Capable	.44
$R_c = .64$	Bossy	.41	Bad judge of personality	.37
$\Lambda = .0185$	Secretive	.30	Touchy	.32
$\chi^2 = 703.9$	Complaining	.28	Masculine	.23
$df = 625$	Sincere	.28	Stingy	.20
$p < .03$	Untouchy	.24		
Males				
No roots removed	Ungullible	.46	Unprying	.32
$R_c = .67$	Unbossy	.41	Sincere	.25
$\Lambda = .0104$	Influential	.36	Uncritical	.23
$\chi^2 = 786.0$	Independent	.25	Greedy	.23
$df = 729$	Secretive	.24	Bossy	.20
$p < .14$				

men, the outcome is more discouraging. The maximal canonical correlation of .67 is not significant ($p < .14$). Table 5 presents the traits contributing the largest weights to each of these canonical correlations.

If we are determined to attempt to interpret the highest correlation in the matrix no matter where it occurs, another approach is suggested. How large does the largest correlation in a matrix have to be so that it is of a magnitude that would only appear once in 100 28 × 28 matrices in which all values were sampling variations on a true r of zero? This seems an appropriate version of the $p < .01$ level. To estimate it, the z value corresponding to $1/(100 \times 28 \times 28)$ or $p = .000012755$ (two-tailed test) has been determined from Fisher and Yates (1953, p. 59, Table VIII_1). This value is 4.37. Analogously, the $p < .05$ value is based upon $1/(20 \times 28 \times 28)$ or $p = .000063775$, $z = 4.01$. Again these values may be judged too conservative if the lack of independence among the 784 values does in fact reduce the sampling variation below that assumed in the formula. Nonetheless, the values resulting are judged to be useful reference points and are presented in the third section of Table 4. Reference back to Table 3 indicates that some of the contrast values exceed these exacting standards. This is particularly clear for gullible, which with −.26 for the combined population has a t value of 5.51, actually significant beyond the whole-matrix $p < .001$ level ($p < .0000012755$, $t = 4.85$). Using these standards, none of the off-diagonal values quite reaches the $p < .05$ ($p < .000063775$) level when the men and women are treated separately. However, for the Peer Attribution × Reputation matrices, a number of values come close and are gratifyingly parallel for men and women. A number of these reach the $p < .05$ and $p < .01$ levels when the two populations are combined and these are listed in Table 6.

Complementary and Reactivity Interpretations of Highest Off-Diagonal Values. Tables 6 and 7 present the most promising off-diagonal projection values when the men and women are pooled by averaging. We have, of course, examined the separate matrices with opportunistic hopefulness and feel that no treasures are lost by this process. Rather, much uninterpretable noise is eliminated. The pooling seems justified not only by the great need for a large N to establish such weak relations (gratified by an increase in the number of statistically significant values when the

TABLE 6

REPUTATION × PEER ATTRIBUTION OFF-DIAGONAL PROJECTION CORRELATIONS

Those with a reputation for being:	Tend to see peers as:	
	Whole matrix $p < .05$ (single value $p < .000063775$)	Single value $p < .01$
Anxious		Suspicious (.13) (comp.) Strict (.13) (comp.) Secretive (.12) (comp.)
Bossy	Gullible (.19) (comp.)	
Capable		Gullible (.16) (react.)
Complaining		
Conceited		Gullible (.17) (comp.) Dependent (.14) (comp.) (Not)Unscrupulous (−.14) (?)
Critical	Gullible (.21) (comp.)	
Dependent	(Not)Gullible (−.20) (?)	Unscrupulous (.12) (react.)
Effeminate		
Greedy		(Not)Suspicious(−.15) (?) (Not)Capable (−.12) (?)
Gullible		Unappreciative(.14) (?) Strict(.12) (?)
Hostile		(Not)Influential(−.12) (?)
Influential	Gullible (.20) (react.)	Dependent(.12) (react.)
Insincere		Gullible(.15) (comp.) (Not)Unappreciative (−.15) (comp.) (Not)Influential(−.13) (comp.) Hostile(.12) (react.)
Obsequious		Secretive(.14) (react.) (Not)Unscrupulous (−.13) (?)

Table 6—Continued

Those with a reputation for being:	Tend to see peers as:	
	Whole matrix $p < .05$ (single value $p < .000063775$)	Single value $p < .01$
Orderly		(Not)Suspicious (−.12) (?)
Prying		Greedy (.14)(?)
Scheming		Gullible(.16) (comp.)
Secretive		
Sex interest		
Stingy		Unappreciative (.12) (react. or comp.)
Strict		(Not)Prying (−.14) (?)
Stubborn		Gullible(.12) (?)
Suspicious		(Not)Influential(−.13) (?)
Touchy		Insincere(.12) (comp.)
Unappreciative		
Unscrupulous		
Good judge of personality	Gullible(.21) (?)	Dependent (.16) (?) (Not)Influential(−.13) (?) (Not)Obsequious(−.12) (?)
(General unfavorability)		

Note.—Abbreviations are: comp.—complimentary; react.—reactivity.

populations are pooled) but also by the fact that none of the theoretical considerations implies different processes for men and women. Because the whole-matrix values as computed are probably too conservative, these tables also present those values which reach the ordinary $p < .01$ level. In the first parenthesis following the trait name is indicated the magnitude of the correlation. In the second parenthesis is indicated the form or projective relationship judged most plausibly to explain the outcome. Since all of these are off-diagonal relationships, similarity and contrast have been excluded as explanations, even where, as discussed later, they might be plausibly rationalized. Where neither of the two off-diagonal hypotheses (complementary pro-

TABLE 7

REPUTATION × PHOTO ATTRIBUTION OFF-DIAGONAL PROJECTION CORRELATIONS

Those with a reputation for being:	Tend to see photo others as:	
	Whole matrix $p < .05$ (single value $p < .0000635$)	Single value $p < .01$
Anxious		
Bossy		
Capable		
Complaining		
Conceited		
Critical		
Dependent		(Not)Conceited (−.12) (?)
Effeminate		
Greedy		(Not)Capable (−.14) (?) (Not)Influential (−.13) (?) (Not)Obsequious (−.12) (?) (Not)Scheming (−.12) (?)
Gullible		
Hostile		
Influential		
Insincere		(Not)Influential (−.13) (?)
Obsequious		Bossy (.14) (?)
Orderly		
Prying		
Scheming		
Secretive		(Not)Greedy (−.13) (?)
Sex Interest		
Stingy		
Strict		Stubborn(.13) (?)
Stubborn		
Suspicious		
Touchy		Anxious(.12) (?)
Unappreciative		
Unscrupulous		
Good judge of personality		
General unfavorability		Strict(.14) (?) Stubborn(.13) (?) Stingy(.12) (?)

jection or reactivity) seemed reasonable explanations, a question mark has been used.

The interpretation of the relationships in terms of the two mechanisms offers so many opportunities for intricate rationalization that these interpretations are merely suggestive, and will not be defended in detail. The conception of these mechanisms, however, will be clarified by illustrating the way in which they have been employed. For complementary projection we ask the question: "Would having this misconception about other persons lead one to behave in this way?" For reactivity we ask: "Does this person's distinctive behavior pattern bring out in others the kind of behavior he is prone to report seeing in them?" Anxious persons tend to see others as suspicious, strict, and secretive. Misperceiving others in this way might tend to make one behave anxiously (complementary projection). But what kind of behavior does an anxious person tend to elicit from others? Were it to seem plausible, as it does not to us, that anxious behavior elicits from others suspiciousness, strictness, and secretiveness, reactivity would have been invoked as the mechanism. Bossy and scheming persons tend to see others as gullible. Misperceiving others as gullible might very plausibly lead one to behave in a scheming exploitative fashion (complementary projection). On the other hand, bossy and scheming behavior would most plausibly elicit resentment from others, not gullible acceptance; thus, reactivity is rejected as a plausible rationalization. Stingy persons tend to see others as unappreciative. This makes good reactivity sense, since the stingy person is reacted to without appreciation even when the stingy person feels he is being generous; thus, stinginess elicits unappreciativeness in others. On the other hand, if upon every occasion for sharing, a person misperceived that the potential recipients would be unappreciative, the sharing would not take place, and the resultant behavior would be stingy. Thus for this relationship, both complementary projection and reactivity are judged very plausible. Capable and influential persons tend to see others as gullible. Since persons who are competent and socially effective would tend to elicit more compliant behavior from others, and hence report more in others, reactivity becomes a plausible mechanism. On the other hand, it does not seem plausible that misperceiving others as gullible would lead to behavior judged capable and influential, so complementary projection is ruled out.

For the few Photo-Attribution pairs listed in Table 5, reactivity is not directly available as an explanatory mechanism. And while complementary projection would be appropriate, it does not seem plausible in the instances listed.

A Contrast Interpretation of the Major Off-Diagonal Values. On an item-by-item basis, the mechanisms of reactivity and complementary projection offer plausible explanations of most of the relationships found, but they do not provide a satisfying explanation of the total pattern of relationships. The most striking fact of Table 6 is the ubiquitous presence of gullible. Not only does it account for all values above .19, but it also provides 5 of the 31 items in the right-hand column; and these have the higher values. Gullible also appeared as the most strongly projected trait in Table 3, and that value of −.26, if repeated in Table 6, would be clearly the largest. In Table 4, the tendency to see others as gullible has been interpreted in some instances as complementary projection and in others as reactivity; the use of such different mechanisms to explain such a similar outcome seems unsatisfactory. Further, two of the high values have gone unrationalized.

Our proposed single-process interpretation is to interpret all of these as manifestations of contrast. Among the Reputation values a factor analysis (not presented in this report) shows a strong general factor of assertiveness-passivity. At the assertive pole lie both good and bad traits: critical, bossy, capable, influential, good judge of personality, and to some extent scheming and stubborn. At the passive pole lie dependent and gullible. Persons who by Reputation are high on this factor tend in contrast to see others as low on it. But the factor is not conspicuous among the Peer Attributions, and gullible seems to be the trait primarily used to express the factor in Attribution. At a lower level, some of the other traits from this cluster do appear as Attributions to add plausibility to the interpretation. Combining data from Tables 3 and 6, it can be seen that dependent persons tend to view others as not dependent (−.12), but still more as not gullible (−.20); and that influential persons tend to see others as not influential (−.14), but still more as gullible (.20). In Table 6, gullible most often appears with traits within this same cluster. Other manifestations of contrast within the assertiveness-passivity cluster are the attributions of dependence by persons reputed to be conceited, influential, and good judges of personality.

This interpretation places contrast and complementary projection as rival mechanisms explaining some of the same specific values in Table 6. In these cases, it is more supporting of contrast, as theoretically developed, that the relationships fail to appear in the Photo Attributions. In the other instances, where contrast is rivaling reactivity, it is congruent with both theories that the relationship only appears in the judgments of persons with whom one is well acquainted.

Inspection of the variables with high loadings on the related canonical variates presented in Table 5, seems to add support to the interpretation of Reputation-Attribution relations as mainly reflecting contrast. As indicated, canonical correlation represents a combination and weighting which maximizes the correlation between the derived canonical variates. Thus, while providing a test of the significance of a matrix, it does not evaluate the relative explanatory power of the several projective mechanisms. Nevertheless, contrast appears to be particularly evident among traits making the largest contribution to the first canonical correlation for females. The variables with the highest weights among Reputations are orderly, scheming, bad judge of personality, and uncritical. Among Attributions, unorderly, unscheming, influential, and independent are most prominent. In view of the high correlations between some traits within Reputations, contrast appears supported even in the cross-trait combinations of variables contributing to this first canonical correlation.

For instance, among Reputations, there is a .75 correlation between bad judge of personality and uninfluential. Likewise, dependent and uncritical are also highly correlated ($r = .58$). If one is intent on discovering support for other mechanisms, some of the traits with high loadings on the other derived canonical variates can of course be paired so as to lend support to complementary projection and reactivity. Nevertheless, we interpret the canonical correlation outcomes as not essentially contradicting our single-process, contrast interpretation of the data.

DISCUSSION

Because of the number and complexity of the analyses, the basic discussions have been placed with the presentation of results in the preceding sections. Here it will be appropriate to review the more general conclusions.

Similarity projection has not been found although it has been put to more exhaustive tests than in any previous study. The occurrence of similarity projection as a function of insight has been carefully examined, with still no evidence of similarity projection. This negative conclusion is consistent with previous studies (excluding those defining the projector's personality in terms of his self-descriptions) when specifiable artifacts in scoring procedures are avoided. Since similarity projection in one form or another is the most common understanding of projection, and since many have expected to find such projection in the trait-attribution setting, this consistently negative outcome deserves attention.

What has been found instead is a significant *contrast* relationship, for some traits, between Reputation and Attribution. This too confirms previous findings, and although the magnitudes of the relations are low, the significance levels are high and the replication across sexes reassuring. These relationships have been called contrast projection in the earlier literature and may therefore be so designated here. However, there are reasons for suspecting that they do not belong to the concept of projection as ordinarily conceived. First of all, our available theory predicting such relationships generates them by assuming that the perceiver accurately notes the direction of differences between himself and others. But in reporting these differences he uses his own behavior as a neutral point, in a fashion analogous to adaptation-level theory, and thereby mislabels the true difference. This dependence of the correlation upon observation of the true characteristics of the persons now being judged sets the concept apart from projection concepts in which the projected content reflects the perceiver's own personality and past environmental influences, wrongly attributed to present persons. This interpretation of the contrast relationship is supported by the fact that it primarily appears when the judgments are of familiar others, but does not appear when the judgments are purely projective attributions of traits to unknown persons represented in the photographs.

The type of projection that our initial theoretical analysis most strongly supported was *complementary projection.* While this mechanism effectively rationalized a few significant projection correlations, since these few are also plausibly explainable in terms of contrast, the economical interpretation is that this type of projection also failed to appear. The only other mechanism with any indication of usefulness in explaining projection correlations was *reactivity*. Here too, a contrast interpretation is also available. Further, the mechanism is again not projective, in that it assumes a report on the actual characteristics of the persons judged.

Perhaps then, we should conclude that we have found no evidence of projection of any kind. If so, this suggests that psychology should restrict the generality of its concept of projection so that it does not apply to a normal person's perception of others in general. If we abandon projection as a general mechanism, what are the more restricted circumstances under which it might still be found? It may be restricted to guilt and wish rather than to more generalized traits. Another possibility is that projection may occur on different traits for different persons. Each person's projections would thus be related to idiosyncratic features of his own past experience. A further limitation is that individuals may project different and quite specific contents onto different types of targets. The targets within each type would be selected on the basis of high similarity in appearance, personality, or role. Thus, frightened boys may see other boys as frightened, but may see men as frightening (Feshbach & Feshbach, 1963). Here it would be assumed that all targets for a particular projective content would be similar in some sense to the person toward whom the projection (or true attribution) was initially learned. Focal projections such as these appear to constitute the core of instances which give clinicians high confidence in the mechanism.

If the operation of the projective mechanism were restricted in any of these ways, the question would then be whether projection could ever be adequately verified or whether such rival hypotheses as complementary and similarity projection could be put to test. One approach might aim at collecting from cooperating clinicians instances in which psychotherapy patients have reacted to new acquaintances with vivid reactions, with personality attributions or accusations which the clinician judges to be of projective content. By combining such instances with the clinician's description of the patient's personality, or still more preferably, with independent clinical assessments of the patient, it might indeed be possible to disentangle complementary and similarity mechanisms.

Even more pessimistic from our standpoint, however, are three other possibilities: that different persons use different projective mechanisms, that both within and across individuals different mechanisms operate for different types of targets, or that within and across individuals different mechanisms operate for different traits. If any of these possibilities were true, the explanatory power of the concept would be severely curtailed. Nevertheless, the practicing clinician would undoubtedly continue to encounter vivid confirmations.

The concept of complementary projection has such magnificant potential as a simplifying construct which unites learning and social cognition theories with clinical practice, that we still have a strong and lingering reluctance to abandon it. Rather than accept any one or any combination of the three possibilities mentioned above, we prefer to think that projection may still have much of the generality which we initially supposed. This belief can perhaps be justified when we re-examine previous studies of projection. When we look for the communality in all the studies which find projection under conditions where a person's characteristics or motivational states are assessed independently from his view of some target, one feature stands out. In all such studies, a strong and usually aversive motive was aroused prior to the attribution task, prior to the measurement of the subject's view of the world (Bramel, 1963; Epstein & Smith, 1957; Feshbach & Feshbach, 1963; Hornberger, 1960; Murray 1933). In retrospect, if we reconsider the conditions under which projection is learned it seems quite reasonable that some form of motivational arousal is a necessary antecedent for eliciting it on subsequent occasions. In the initial conditions under which a person learned to attribute traits with evaluative connotation, motivational arousal was undoubtedly high. Thus, the learned attribution tendency, whether similarity or complementarity, may only generalize to new targets or to new traits when some form of threat is present. As previously pointed out, most of the prior projection studies do not use an experimental design which permits specification of whether any obtained projection supports a similarity or complementary mechanism. In the studies which do allow separation of the two interpretations, complementarity is supported. Hornberger (1960) finds no similarity projection, and though Feshbach and Feshbach (1963) prefer to argue that different projective mechanisms are employed for different targets, their outcome, too, primarily supports complementary projection. Thus, our initial expectations of ubiquitous complementary projection may still be only a slight exaggeration. The complementary mechanism may indeed operate universally and for numerous traits whenever persons are motivationally aroused. Furthermore, the occurrence of complementary projection, or for that matter similarity projection, would not be logically incompatible with the contrast mechanism found in the present study. When the projective targets are known persons from the immediate environment, the contrast judgment mechanism may simply be superimposed.

If this analysis is correct, it suggests a replication of our basic design with only a slight procedural modification. Each person must be made sensitive, worried, or concerned about what kind of person he is. His reputation must become salient and potentially threatening or aversive *before* he is asked how he sees his peers or unknown photo persons. The experiment can be presented as an attempt to see if insight can be gained by having the group discuss, in turn, each individual's personality, stressing in particular his flaws and weaknesses on each of the traits. To add to the potential threat of the situation, the experimenter could state that the results of a previously administered Rorschach test would be disclosed in order to provide some foundation for the discussion of each individual. The character dissection session would supposedly follow the peer and photo rating tasks. By thus insuring that each person's traits are salient and self-threatening, systematic distortions in attributions may more reasonably be expected to appear.

SUMMARY

Personality has been assessed by peer reputation, and projection by the attribution of traits to peers and to unknown photographed persons. No evidence of *similarity* projection has been found, even when insight, measured in three different ways, has been taken into consideration. *Contrast*, predicted from adaptation-level theory, has been found for the traits gullible, secretive, orderly, influential, and suspicious. The predicted instances of *complementary* projection (a form of projection following from cognition and learning theories) were not confirmed in general. Whereas complementary projection and reactivity processes could explain most of the significant off-diagonal projection correlations (i.e., correlations in which the personality trait and the projected trait differ), the more parsimonious interpretation is to regard these also as showing a *contrast* process in which assertive and exploitative persons judge others to be passive and gullible.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice Hall, 1952.

BACON, F. *The physical and metaphysical works of Lord Bacon*. (Trans. by J. Dewey) London: Bohn, 1853.

BLOCK, J. On the number of significant findings to be expected by chance. *Psychometrika*, 1960, **25**, 369–380.

BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **49**, 94–98.

BRAMEL, D. Selection of a target for defensive projection. *J. abnorm. soc. Psychol.*, 1963, **66**, 318–324.

BRUNER, J. S., & POSTMAN, L. An approach to social perception. In W. Dennis (Ed.), *Current trends in social psychology*. Pittsburgh: Univer. Pittsburgh Press, 1948. Pp. 71–118.

CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.

CAMPBELL, D. T. A typology of tests, projective and otherwise. *J. consult. Psychol.*, 1957, **21**, 207–210.

CAMPBELL, D. T. Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 6. *Investigations of man as socius: Their place in psychology and the social sciences*. New York: McGraw-Hill, 1963. Pp. 94–172.

CAMPBELL, D. T., & BURWEN, L. S. Trait judgments from photographs as a projective device. *J. clin. Psychol.*, 1956, **12**, 215–221.

CAMPBELL, D. T., & FISKE, D. F. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81–105.

COOLEY, W. W., & LOHNES, P. R. *Multivariate* *procedures for the behavioral sciences.* New York: Wiley, 1962.

EPSTEIN, S., & SMITH, R. Thematic apperception, Rorschach content, and ratings of sexual attractiveness of women as measures of the sex drive. *J. consult. Psychol.*, 1957, **21**, 473–478.

FESHBACH, S., & FESHBACH, N. Influence of the stimulus object on complementary projection of fear. *J. abnorm. soc. Psychol.*, 1963, **66**, 498–503.

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural, and medical research.* (4th ed.) New York: Hafner, 1953.

FREUD, ANNA. *The ego and the mechanisms of defense.* London: Hogarth Press, 1937.

FREUD, S. *Totem and taboo and other works.* London: Hogarth Press, 1955.

HELSON, H. Adaptation-level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. *Sensory, perceptual, and physiological formulations.* New York: McGraw-Hill, 1959. Pp. 565–621.

HORNBERGER, R. H. The projective effects of fear and sexual arousal on the ratings of pictures. *J. clin. Psychol.*, 1960, **16**, 328–331.

JAMES, W. *The principles of psychology.* New York: Holt, 1890. 2 vols.

LEWIN, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936.

LEWIN, K. *Field theory in the social sciences.* New York: Harper, 1951.

LUBETSKY, J. Assimilative projection as measured by trait attribution. Unpublished doctoral dissertation, Northwestern University, 1960.

MARKS, E. S. Skin color judgments of Negro college students. *J. abnorm. soc. Psychol.*, 1943, **38**, 370–376.

MURPHY, G. *Personality.* New York: Harper, 1947.

MURRAY, H. A. The effect of fear upon the estimates of the maliciousness of other personalities, *J. soc. Psychol.*, 1933, **4**, 310–329.

MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

PENFIELD, W., & RASMUSSEN, T. *The cerebral cortex of man.* New York: Macmillan, 1950.

RICKMAN, J. *Sigmund Freud: A general selection.* London: Hogarth Press, 1937.

ROKEACH, M. Studies in beauty: II. Some determiners of the perception of beauty in women. *J. soc. Psychol.*, 1945, **22**, 155–169.

SEARS, R. R. Experimental studies of projection: I. Attribution of traits. *J. soc. Psychol.*, 1936, **7**, 151–163.

SHERIF, M., & SHERIF, C. W. *An outline of social psychology.* (Rev. ed.) New York: Harper, 1956.

SULLY, J. *Illusions: A psychological study.* New York: Appleton, 1881.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WYLIE, RUTH C. Some relationships between defensiveness and self-concept discrepancies. *J. Pers.*, 1957, **25**, 600–616.

WYLIE, RUTH C. *The self concept: A critical survey of pertinent research literature.* Lincoln: Univer. Nebraska Press, 1961.

ZIMMER, H. Projection of personality characteristics as a function of ambivalence over internalized demands. Unpublished doctoral dissertation, University of Rochester, 1952.

(Received January 17, 1963)

Vol. 78, No. 16 Whole No. 593, 1964

Psychological Monographs: General and Applied

EXTINCTION OF ASSOCIATIONS IN FOUR TRANSFER PARADIGMS

JEAN BARNES MCGOVERN[1]

Northwestern University

3 factors were identified as involved in the production of retroactive inhibition (RI) in immediate, written free recall of 4 transfer paradigms: (a) extinction of first-list S-R or forward associative connections; (b) extinction of first-list R-S or backward associations; and (c) extinction in response-recall, i.e., extinction of associative connections formed between environmental or contextual stimuli and first-list responses. The first and third factors were hypothesized as operating in the A-B, A-C paradigm. The first and second factors were hypothesized as operating in the A-B, A-Br paradigm. The A-B, C-B and A-B, C-D paradigms were included for independent verification of Factors b and c, respectively. Results verified all hypotheses concerning the presence of extinctionlike factors operating to produce RI in the 4 paradigms. Results indicate that paired-associate recall may be divided into 2 stages: a response-recall phase and an association phase. A loss in recall apparently may occur in 1 phase independent of loss in the other phase. Loss in recall in the associative phase may be produced by Factors a and/or b listed above. Losses in in response-recall were attributed to Factor c.

THIS study deals with transfer theory and its implications for immediate recall. More specifically, it is an attempt to extend the implications of a recent set of experimental results (Barnes & Underwood, 1959) to a number of transfer situations.

One of the conclusions arrived at in the Barnes-Underwood study was that an extinction of first-list associations occurs during second-list learning in the A-B, A-C[2] negative transfer paradigm producing a decrement in the recall of first-list responses. The present experiment has been designed to extend this interpretation by analogy to three other transfer situations. The strategy employed is to analyze these other experimental arrangements in a search for the presence of A-B, A-C (extinctive) relationships between first- and second-list associations. Such an analysis then makes it possible to predict and understand the mechanism through which each of these situations leads to the production of retroactive inhibition (RI).

The Barnes-Underwood study was designed to determine the course of retention of first-list associations during the learning of a second list. Only the part of the study involving the A-B, A-C paradigm need concern us here. In this study subjects learned the first list to a criterion of one perfect trial. One minute later they began a series of 1, 5, 10, or 20 trials on the second list. Finally they took an immediate written test of free recall in which they were instructed to write *both* first- and second-list responses as they came to mind and to pair them with the appropriate stimuli on a mimeographed sheet.

---

[1] This is a revision of the author's doctoral dissertation (Jean M. Barnes, 1960) submitted to the Graduate School at Northwestern University. The research was conducted while the author was a National Science Foundation Cooperative Fellow.

The author is indebted to B. J. Underwood for his patient guidance in all phases of this research. Sincere thanks are given to Nan E. McGehee and J. Richardson for their helpful suggestions.

[2] Conventional notation is employed here. A pair of letters (e.g., A-B) stands for the stimuli (A) and responses (B) in a paired-associate list. Two such pairs (e.g., A-B, A-C) identify two lists and the relationships between them. Thus the A-B, A-C paradigm is one in which the subject learns in succession two lists in which stimuli are identical (A), but responses differ (B in List 1; C in List 2).

Analysis of results for the four groups indicated that there was a gradual decrease in recall of first-list (original learning or OL) responses as the number of trials increased on the second list (interpolated learning or IL) so that with 20 trials on the second list, only about 50% of first-list responses were recalled. Results in a control group indicated that this loss could not be attributed to normal forgetting. Comparison of results from two methods of measuring first-list recall showed that if first-list responses were recalled, the subject knew which stimulus they were paired with in OL and that they were first-list responses. In addition, the subjects' verbal reports and the order of response recall both showed that mediation of second-list responses by first-list associations was not involved. It was concluded that extinction or unlearning of first-list associations during second-list learning is responsible for almost all RI obtained in the A-B, A-C paradigm.

The conclusion of Barnes and Underwood which attributes retroaction to extinction or unlearning of first-list responses in the A-B, A-C paradigm remains somewhat tentative for two reasons. First, instructions to the subjects stressed the pairing of recalled responses with *appropriate* stimuli. Subjects were never told to guess if they were not certain of the stimuli with which recalled responses were associated. Thus, the subjects may have recalled all of the first-list responses, but may have written down only those responses for which they were reasonably certain of first-list stimuli. Hence, first-list responses may not have been lost from the subjects' repertoires, but merely withheld because the subjects were not certain of first-list associations. Second, the control group employed was appropriate only for normal, uncontrolled forgetting. A question might be raised as to whether or not the learning of a new set of responses, per se, in the second list was responsible for loss of first-list responses in recall or whether extinction of specific first-list S-R associative connections alone was involved. A check on this possibility could be made by employing as an additional control the A-B, C-D paradigm (stimuli different; responses different). The present experiment attempted to deal with these two difficulties by (*a*) stressing guessing in the instructions for recall and (*b*) by employing an A-B, C-D condition as a part of the experimental design.

The results of the Barnes-Underwood study also raise two additional questions. First, does an A-B, A-C relation between sets of associative connections always imply an extinctive relationship between two lists? Second, how general is the loss of first-list responses from the subjects' repertoires; are the responses lost only in terms of their specific first-list associative connections, or are they lost as responses to any stimulating situation for which they are appropriate? These questions identify further problems investigated in this study.

Analysis of the Four Transfer Paradigms

The specific methods employed in this study were designed to examine the applicability of the extinction or unlearning theory to four different experimental paradigms employed for the investigation of transfer: (*a*) the A-B, A-Br paradigm in which the same stimulus terms (A) and response terms (B) are used in original and interpolated learning, the terms being re-paired in the second list; (*b*) the A-B, C-B paradigm in which the stimuli in the two lists are different, but the responses are identical; (*c*) the A-B, C-D paradigm where stimuli and responses in the two lists are both different; and (*d*) the A-B, A-C paradigm where stimuli are the same and the responses differ.

The last of these, the A-B, A-C paradigm ("new responses to old stimuli") defines the relationship of second learning to first learning traditionally assumed to produce negative transfer. It also defines the extinctive relationship as that expression is used in this monograph. In attempting to extend the theory of unlearning to the other transfer situations, it will be convenient to approach the problem by asking for each of the other paradigms whether the extinctive relationship (A-B, A-C) exists in each and should, thus, tend to produce retroactive inhibition. The considerations are somewhat different depending upon whether retention is under conditions of modified free recall (MFR) or whether responses are supplied. We shall deal with the former situation first.

The A-B, A-Br Paradigm

Historically, the A-B, A-C paradigm has been considered the optimal condition for producing negative transfer and has been the paradigm around which much RI theory has been built (Osgood, 1953). A wealth of evidence (Besch & Reynolds, 1958; Porter & Duncan, 1953; Twedt & Underwood, 1959), however, indicates that the A-B, A-Br paradigm far exceeds the A-B, A-C paradigm in the production of overt intrusions and negative transfer. Only one study (Mandler & Heineman, 1956) reports positive transfer for the A-B, A-Br paradigm. This study is unique in that consonant syllables served as responses rather than the common words used in the other studies. In addition, high degrees of overlearning of the first list were produced. On these grounds it seems very likely that the large amount of response integration which must have been required by the consonant syllables and which must have accompanied overlearning was responsible for the positive transfer.

The fact that the A-B, A-Br paradigm is usually particularly effective in producing negative transfer suggests that it will be fruitful to begin by asking whether the A-B, A-C extinctive relationship applies to it. An analysis of the interlist relationships resulting from the re-pairing of first-list items in the second list of the A-B, A-Br paradigm points to two possible sources of such extinction: (*a*) In terms of S-R associations, an A-B, A-C relationship exists between the two lists of the A-B, A-Br paradigm because each stimulus of the first list is re-paired with a different first-list response in the second list. (*b*) In terms of backward or R-S associations, an A-B, A-C relationship also exists between the two lists of the A-B, A-Br paradigm because each first-list response is re-paired with a different first-list stimulus in the second list. Thus, to the extent that backward associations develop in paired-associate learning, and to the extent that they are involved in recall, extinction of R-S associations provides a second source of retroaction in the A-B, A-Br paradigm.

These two extinction factors may be described more exactly for the A-B, A-Br condition as follows. Let S_1, R_1 and S_2, R_2 represent two stimulus and response pairs in the A-B or first list of the paradigm. According to the notions presented above, two associative connections are formed between S_1 and R_1 in first-list learning, a forward association, $S_1 \rightarrow R_1$, and a backward or R-S association, $R_1 \rightarrow S_1$. Similar associative connections are established between S_2 and R_2. Suppose that in the second or A-Br list S_2 and R_1 are paired together and S_1 and R_2 are paired together. Now the forward associations that the subject must learn are $S_2 \rightarrow R_1$ and $S_1 \rightarrow R_2$. Similarly, the backward associations that the subject may learn in the A-Br list are $R_1 \rightarrow S_2$ and $R_2 \rightarrow S_1$. Both of these relationships satisfy the defining conditions for an A-B, A-C relationship in terms of S-R (and R-S) associations: different responses are paired with the same stimulus.

The A-B, C-B Paradigm

The analysis just applied to the A-B, A-Br paradigm holds for the A-B, C-B paradigm (stimuli different; responses identical) only in the case of backward associations. In orginal learning the subject masters a set of forward associations A-B and simultaneously the backward associations, B-A. In interpolated learning the new forward association (C-B) does not involve the extinctive relationship, but the backward association (B-C) does. Thus, again, to the extent that backward associations are important in recall, the A-B, C-B paradigm contains the potential for producing retroactive inhibition.

The possible extinction of first-list backward associations, and their role in immediate free recall, can be assessed by a comparison of the A-B, C-D and A-B, C-B paradigms if experimental conditions are arranged with care. Theoretically, the forward (S-R) associations in the A-B, C-B paradigm have zero associative transfer value when intralist response similarity is low (Young & Underwood, 1954) and when response integration is not involved (Mandler, 1954). The only vehicle for the conveyance of transfer is provided by the backward (R-S) associations which, as we have seen, have an A-B, A-C relationship. Hence, with the proper experimental conditions, extinction of first-list backward associations should occur in the A-B, C-B paradigm. In the A-B, C-D paradigm, all associations have zero transfer value. Thus, by comparison with this latter condition, the A-B, C-B arrangement should favor the development of retroaction.

Thus far, only one experiment (Twedt & Underwood, 1959) has demonstrated negative transfer in the A-B, C-B paradigm as compared to the A-B, C-D control condition. At least two studies (Bugelski & Cadwallader, 1956; Hamilton, 1943) have found slight positive transfer and retroactive facilitation for the A-B, C-B paradigm under these circumstances. These results, which contradict those predicted by the theory, may be explainable in terms of a lack of opportunity for the formation and extinction of R-S associations. The lists employed in these studies contained common words as responses and nonsense figures as stimuli. Such materials would probably not permit R-S associations to develop readily; hence, an A-B, A-C extinction process could not operate in terms of backward associations. In the Twedt-Underwood study, on the other hand, common adjectives served both as stimuli and responses, thus, presenting optimal conditions in terms of meaningfulness (Jantz & Underwood, 1958) for the development of R-S associations. Thus, the Twedt-Underwood study was the only one which met the conditions required for the production of negative transfer through the extinction of backward associations. The results of this study suggest that the prediction of negative transfer on the basis of the extinction of R-S associations in the A-B, A-Br and A-B, C-B paradigms have some basis in fact.

The A-B, A-C Paradigm

Analysis of the A-B, A-Br paradigm revealed two possible sources of negative transfer in the form of demonstrable A-B, A-C relationships. These were the extinction or unlearning of forward (S-R) associations and backward (R-S) associations. In the A-B, A-C relationship it is clear that the first of these sources of decrement applies, but that the second does not. Further analysis of the A-B, A-C paradigm, however, suggests a third negative factor not present in the A-B, A-Br or A-B, C-B paradigms. This third factor will be discussed next.

Logically and empirically, paired-associate learning can be divided into two successive phases or stages: the response-recall phase and the associative phase (Underwood, Runquist, & Schulz, 1959). In the first, response-recall phase, such learning takes place as is necessary for the subject to discriminate the response required in learning from his repertoire of other possible responses and to make the appropriate responses readily available for recall. Following this first step, the association phase in which the subject connects stimuli and responses begins. Evidence indicates that learning is facilitated when the subject is taught responses prior to learning a list of paired-associates. Obviously such first-phase learning would facilitate second-list learning in the A-B, A-Br and A-B, C-B transfer situations, where the responses learned in List 1 are the same as those in List 2 and no second response-recall phase of learning is required. As Twedt and Underwood have indicated, any negative transfer which shows up in the second list of either of these two paradigms does so in spite of the facilitation of prior learning of the second-list responses. In the A-B, A-C paradigm, on the other hand, no such prior learning of second-list responses has occurred. To the extent that such learning facilitates recall, this facilitation is lacking in the A-B, A-C paradigm.

The A-B, C-D Paradigm

The identification in the previous section of a source of retroaction associated with the response-recall phase of paired-associate learning inevitably leads to questions as to the more exact nature of the process by which the decrement is produced. The A-B, A-C paradigm does not lend itself well to dealing with such questions because it contains a presumably more potent source of retroaction in the extinctive relationship between forward associations. This relationship does not exist in the case of the A-B, C-D paradigm; however, any retroactive influence assignable to the response-recall phase may be demonstrated independently of any other extinctive processes by a comparison of recall in the A-B, C-D paradigm with recall under control conditions, without interpolated learning. Bugelski and Cadwallader (1956) and Gibson (cited in Hamilton, 1943) present evidence which points to a loss in first-list recall in the A-B, C-D paradigm when it is compared with such a control.

A question must now be raised as to whether the considerations outlined above can account for retroaction in experiments of this type or whether additional factors must be taken into account. Several possibilities of the latter sort suggest themselves. The first is that retroaction might result from a general "set" which develops in the first trial of second-list learning. Seeing that a new set of responses is used in the second list, the subject might instruct himself not to give first-list responses. Immediate recall of first-list responses following IL might then be retarded by the persistence of this set despite the experimenter's instructions to the contrary. A number of studies have demonstrated both facilitatory and inhibitory effects of set on recall (e.g., Jenkins & Postman, 1949; Lester, 1932; Postman & Postman, 1948; Thune, 1958; Waters & Peel, 1935; Whitely, 1927). That such an explanation is untenable as a complete account of retroaction in the A-B, C-D experiment, however, is implied by results obtained in the Twedt-Underwood study. Two procedures were used with four transfer paradigms (A-B, A-C; A-B, A-Br; A-B, C-B; A-B, C-D) in this study: (*a*) unmixed lists (UL) in which the transfer relationships between all pairs of items in the two lists were the same, and (*b*) mixed lists (ML) in which four subgroups of items produced the four different transfer paradigms within lists. In the UL procedure, the development of a set not to give first-list responses during second-list learning was appropriate for the A-B, A-C and A-B, C-D paradigms, since the response terms in the second list were, in each case, new. In the ML procedure, such self-instruction would be appropriate only for two subgroup items. A *general* "set" of not responding with first-list responses in second-list learning would not be appropriate in the ML procedure, and presumably it would not develop. If this is the case, overt intrusion and negative transfer should be greater for the A-B, A-C and A-B, C-D paradigms in the ML procedure than in the UL procedure. However, differences between UL and ML procedures were slight and in a direction opposite to that predicted for the role of set, i.e., negative transfer was lower in the ML condition than in the UL condition for all transfer paradigms.

The unlearning theory under test in this study suggests a different mechanism leading to the production of retroactive inhibition and, in the process of doing so, also suggests something about the nature of the response-recall phase of paired-associate learning if we assume that such learning is a form of associative learning in which the important stimuli are context stimuli, that is, stimuli emanating from the room, the experimental equipment, etc.[3] The learning of two different sets of responses in the A-B, C-D experiment thus establishes an A-B, A-C relationship where the stimuli (A) involved are environmental stimuli. For this reason, extinction of associative connections between context cues and responses may occur in the response-recall phase of an experiment whenever two different sets of responses have been learned, i.e., in the A-B, A-C and A-B, C-D paradigms. Such extinction should not be involved where response systems are identical, i.e., in the A-B, A-Br and A-B, C-B paradigms. According to the above reasoning, the A-B, A-C extinction relationship involving environmental stimuli should be abolished when IL occurs in a different experimental situation. Consequently, RI should be reduced if IL takes place in an environmental situation different from that in which OL and RL take place. At least two studies (Bilodeau & Schlosberg, 1951; Greenspoon & Ranyard, 1957) have obtained results which support this prediction. RI was significantly reduced if IL took place in a different experimental room as compared to a condition in which OL and IL took place in the same experimental room. These results correspond to interpretations emphasizing the importance of context stimuli in the control of paired-associate learning.

---

[3] It is likely, of course, that there is more to it than this. The response-recall phase quite probably involves at least two mechanisms: response integration and association to context cues. In this study the emphasis has been almost entirely on the latter process.

Recall When First-List Responses Are Supplied

The primary function of context cues appears to be that of prompting the recall of responses as a group, as seems to have been demonstrated by certain results obtained in the Barnes-Underwood study. In this study, employing the A-B, A-C paradigm, it was found that, if a subject recalled a first-list response, he could usually pair it with the correct stimulus in written recall and identify it correctly as being a first-list response. This raises the question whether the subject could correctly pair *all* first-list responses with first-list stimuli in immediate written recall if first-list responses were supplied. In a sense these responses are provided in the A-B, A-Br and A-B, C-B paradigms. Response-recall, independent of associative connections with list stimuli, should be no problem for the subject since the same set of responses is employed in Lists 1 and 2. In order to investigate such matters in this experiment, half of the subjects in each transfer paradigm were provided with the List 1 responses during recall with instructions to match them to appropriate first-list stimuli. This procedure should eliminate the effects of any extinction process which made the responses unavailable. According to our theory such extinction takes place only in the A-B, A-C and A-B, C-D paradigms where the extinction of first-list responses to context cues occurs. Retention in the A-B, C-D paradigm, then, should be perfect when the subject is provided with the set of first-list responses in recall; and retention in the A-B, A-C paradigm should be improved as compared to the usual conditions of free recall in which the subject is presented with first-list stimuli alone. Probably, recall in these paradigms would also be enhanced because providing the subject with correct responses also provides him with an opportunity to guess. Retention in the A-B, A-Br and A-B, C-B paradigms should be relatively unaffected by the procedure of giving the subject a set of first-list responses in recall since essentially the same conditions obtain in the usual recall situation due to the nature of the paradigm. Again, any remaining confusion in recall would reflect extinction of S-R and/or R-S associative connections in the first list.

Generality of Extinction

We come finally to the question of the generality of the effects produced in the RI experiment. One way to approach this issue is to ask the following question: In the retroaction experiment, is the extinguished response lost from the subject's repertoire only as a reaction to the specific first-list stimulus, or is a more general type of inhibition developed so that the item cannot be used as a response in another stimulus situation where it is appropriate, even when the contributions of stimulus generalization (Rothkopf, 1957; Saltz, 1953a, 1953b) have been eliminated? To investigate this issue one might examine the frequency of usage of "extinguished" responses in a "free-association" situation where they would be appropriate. If general inhibition of a particular response develops during the extinction of a specific

S-R connection in IL, then the frequency of occurrence of this response should be reduced, following IL, in a standard free-association experiment.

Summary of Predictions

The theory outlined above makes several predictions for the outcome of this experiment which studied recall of A-B associations under four experimental paradigms: A-B, A-C; A-B, A-Br; A-B, C-B; and A-B, C-D. It will be useful, at this point, to present the predicted outcomes in summary form.

Modified Free Recall: First-List Responses Not Provided

Three sources of negative transfer and RI have been identified: (*a*) extinction of first-list S-R or forward associations; (*b*) extinction of first-list R-S or backward associations; (*c*) extinction of associations between environmental stimuli and first-list responses. At least one of these factors is assumed to operate in each of the four transfer paradigms investigated in the present study. Thus RI is expected to occur in all transfer conditions by comparison with a control condition without IL.

As we have seen, the extinctive (A-B, A-C) relationship applies to different associations under different experimental paradigms as follows: A-B, C-D—context associations; A-B, C-B—backward associations; A-B, A-Br—forward and backward associations; A-B, A-C—forward and context associations. On the basis of such an analysis, unlearning theory predicts that recall will be poorest in the A-B, A-Br and A-B, A-C conditions since these two conditions each have two negative factors operating to reduce recall. Predicting from earlier transfer studies which demonstrate a greater amount of negative transfer in the A-B, A-Br than in the A-B, A-C condition, recall should be poorest in the A-B, A-Br condition.[4] Recall in the A-B, C-B and A-B, C-D conditions should exceed that found in the A-B, A-C and A-B, A-Br conditions because only one negative factor has been identified in the former conditions. Of these two conditions, it is expected that recall will be higher in the A-B, C-D condition than in the A-B, C-B condition because previous studies of transfer have shown poorer second-list performance in the A-B, C-B condition than in the A-B, C-D paradigm (Twedt & Underwood, 1959).

List Recall: First-List Responses Supplied

Providing the subject with a list of correct responses and requiring him only to produce the correct associations should eliminate retroactive influences assignable to the response-recall phase of retention. According to the theory developed above, this means that it should eliminate retroaction produced by an A-B, A-C relationship involving context cues. On this basis it is to be expected that recall will improve under this condition in the A-B, A-C paradigm since one of two retroactive influences is eliminated. In the A-B, C-D condition, recall should be perfect since extinction of associative connections between context stimuli and responses is the only negative factor postulated for this condition. Recall in the A-B, A-Br and A-B, C-B paradigms should be relatively unaffected by supplying the subjects with first-list responses since the same set of responses is used in first- and second-list learning. In these two paradigms, all responses are supplied in recall by the nature of the experimental procedure.

METHOD

The subjects learned two lists of paired-associates of one of the following relations: A-B, A-C; A-B, A-Br; A-B, C-B; A-B, C-D. The two lists were learned during one experimental session with a 1-minute rest interval between them. The lists were typed on white vellum tape and presented, employing the anticipation method, at a 2:2-second rate on a Hull-type memory drum. The intertrial interval was 4.0 seconds. The criterion of learning for the first list was one errorless trial. The second list was presented for 15 anticipation trials. Following practice on List 2, half of the subjects were given a written free-recall test (MFR) in which they were given first-list stimuli and asked to write down first-list responses after receiving 15 anticipation trials on the second list.

---

[4] This analysis treats the several types of associations as if they were of equal strength and therefore equally easy to extinguish. It seems fairly certain that such an assumption is an oversimplification; but the relative strengths of the various associations must remain a matter for experimental investigation in the future.

The other half of the subjects in each condition were also given a list of the responses learned in List 1 and were instructed to pair them with first-list stimuli. These subjects formed the *list-recall* subgroups of the four transfer paradigms. The remaining subjects in each transfer condition formed the *free-recall* subgroups. The two conditions of recall in the four transfer paradigms form eight independent groups. The four transfer groups employed in the present experiment will be identified, hereafter, by their second lists. Thus, the group learning the two lists of an A-B, A-C paradigm will be identified as the A-C condition; the group learning two lists of an A-B, A-Br relation will be identified as the A-Br condition; etc.

In addition to these eight experimental groups, a control group was run as a check on the possibility that uncontrolled "forgetting," rather than the unlearning or extinction introduced by second-list learning, was responsible for any obtained decrement in recall of first-list responses in the transfer conditions. These subjects spent 9 minutes on an unrelated task, 9 minutes being the amount of time taken by the experimental subjects for practice on List 2. All subjects received a free-association task following recall.

Subjects and Design

The 216 elementary psychology students who served as subjects were assigned to one of the nine conditions described above by the following procedure. The nine conditions were arranged in six independent orders. The subjects were then assigned in succession to a condition within one of these six blocks as they appeared at the laboratory. Four subjects were assigned to each condition within each block. Thus there were 24 subjects in each experimental condition. Twelve subjects were dropped and replaced when they admitted to attempting rehearsal of List 1 during the learning of List 2.

Materials

The first list of paired associates was identical for all nine conditions. The stimuli in all lists consisted of nonsense syllables with 53%–73% Glaze association value (Stevens, 1951, p. 542). Intralist similarity of stimuli was kept at a minimum. A given consonant was used but once in the eight syllables of a given list, and three of the five vowels were used twice. In conditions where different stimuli were used in the first and second lists, i.e., in the A-B, C-D and A-B, C-B paradigms, interlist stimulus similarity was kept at a minimum; none of the stimuli in the second list started with a consonant used as the initial letter of any stimulus in the first list. In the A-B, C-B and A-B, A-Br paradigms, the two stimuli associated with a given response had no letters in common. Responses in all lists consisted of common adjectives. Intralist and interlist response similarity were also minimized where appropriate. First-list responses (JOYFUL, EVIL, GREAT, FATIGUED, CALM, FRIGHTENED, UNFILLED, FILTHY) were obtained empirically in a "free-association" task described below. Second-list responses which were not identical to first-list responses were chosen for their apparent lack of meaningful similarity to first-list responses from Haagen's (1949) lists of two-syllable adjectives.

Four orders of presentation of list pairs were used to prevent serial learning of the responses. The starting order of the first list for each group was completely counterbalanced throughout the nine experimental conditions.

Procedure

Standard paired-associate learning instructions were read to the subjects at the beginning of the experimental session. Brief instructions for second-list learning were read to the subjects in the 1-minute interval between List 1 and List 2. The subjects were stopped after 15 anticipation trials on List 2, given a mimeographed list of stimulus terms, and instructed to write the first-list responses they had learned opposite the appropriate stimuli. The instructions emphasized that the subject should guess when he was unsure of a particular association. The stimuli on the mimeographed sheet were presented in the order which the subject would have received them on the trial following the criterial trial in first-list learning. This method of presenting first-list stimuli in recall was expected to maximize recall. As was noted above, half of the subjects in each transfer condition received an index card with the set of words used as first-list responses printed on it. These subjects formed the list-recall subgroups of the transfer conditions. The remaining subjects in each paradigm formed the free-recall subgroups. In all cases, the subjects were given a 2-minute period in which to write.

Control Group

The control condition, the four orders of the A-B list were used as starting orders equally often. Following learning and a 1-minute rest interval, these subjects were given a puzzle to solve in which they were required to remove a pyramid of wooden blocks from one of three pegs and restack them in order on another peg. The subjects spent 9 minutes on the puzzle which is approximately equivalent to the time (9½ minutes) spent by experimental subjects on the second list in the transfer conditions. Following written free recall, these control subjects received the free-association task as in the other eight conditions of this experiment.

Free Association

In the free-association task the subjects were presented with eight adjectives (EMPTY, SOILED, MERRY, WICKED, WEARY, IMMENSE, PEACEFUL, and AFRAID) on a mimeographed sheet. In the spaces provided, the subjects were to write the first three synonyms that they could think of for each word. These words had previously been presented to two groups of experimental psychology students ($N = 40$) who were required to list as many synonyms as possible for each key word in a 20-minute period. The first-list responses which were used in this study were chosen from the responses supplied by these two groups. Each of the response words chosen as a first-list response occurred as a synonym to one of the key words between 30% and 50% of the time. These particular frequencies were chosen because it was felt that they would most clearly reveal any general extinction effects produced by IL. In the free-association task of the present experiment, the subjects were allowed 3 minutes in which to write down the synonyms that they wished to use.

Interrogation

At the end of the experimental session, the subjects were asked a number of questions designed to determine: (*a*) if they made any deliberate attempt to retain first-list responses, (*b*) if they used first-list stimuli or responses to mediate second-list learning, (*c*) if they noticed the relationship between the free-association key words and first-list responses.

RESULTS

Equivalence of Groups

The procedure used to assign the subjects to the eight experimental conditions and the control group was successful in producing groups essentially equivalent in learning ability. Analysis of variance indicated that the groups did not differ significantly in mean number of trials to reach a criterion of one perfect trial on the first list ($F < 1.00$ for the eight experimental groups; $F < 1.00$ for the free-recall subgroups and the control group; $F = 1.44$, $df = 4/115$ for the four list-recall subgroups and the control group). The mean number of trials to criterion for the nine groups was 9.52, with a range in means from 8.29 to 10.67. Inspection of successive criteria curves also indicated that the nine groups did not differ appreciably in first-list learning. To the extent that trials to criterion on the first list is a valid measure of learning ability, the random assignment of subjects to groups was effective in equating groups in learning ability.

Transfer Effects and Intrusions

Before any meaningful statement can be made concerning transfer effects in second-list performance (transfer effects), it must be demonstrated that the four transfer lists which define the transfer paradigms do not differ significantly in difficulty. Following the completion of the present study, four additional independent groups of 12 subjects each were given one of the four transfer lists to learn. Training was continued until the subject reached two criteria: (*a*) a criterion of one perfect trial and (*b*) a criterion of at least 15 anticipation trials. An analysis of variance performed on trials to criterion indicated that the lists did not differ significantly in difficulty ($F < 1.00$). The overall mean of trials to criterion was 12.44; the range for different groups was from 11.50 to 13.33. Inspection of a plot of mean number of correct responses over 15 trials and inspection of successive criteria curves also support the conclusion in favor of equal list difficulty for the four lists which define the four transfer paradigms.

Returning to the experiment proper, an analysis performed on the total number of correct responses in 15 trials for second-list learning revealed no significant difference in results for the free-recall and list-recall subgroups of the four transfer paradigms ($F < 1.00$). This result is in accord with experimental procedure since the recall subgroups within each transfer condition had not been treated differentially up to this point. On the other hand, a significant difference was obtained in a comparison of the four subgroups which differed in terms of transfer paradigm ($F = 13.09$, $df = 3/184$, $p = .01$). Consequently, individual t tests were performed on the differences in total number of correct responses over 15 trials for the four transfer paradigms. For this analysis the scores for the two recall subgroups within each paradigm were combined, making a total of 48 subjects in each transfer group. The results of these t tests are summarized in Table 1. Inspection of Figure 1 indicates that the ordering of the transfer groups remains relatively constant throughout the 15 trials of second-list learning.

The results presented in Table 1 and Figure 1 indicate that the present study

TABLE 1

DIFFERENCES IN MEAN NUMBER OF CORRECT RESPONSES OVER 15 TRIALS FOR SECOND-LIST LEARNING

Paradigm	Mean number of correct responses	Difference		
		A-Br	C-B	C-D
A-C	86.96	5.31	2.42	14.75**
A-Br	81.65	—	7.73*	20.06**
C-B	89.38	—	—	12.33**
C-D	101.71	—	—	—

* $p < .05$, t test.
** $p < .01$, t test.

deals with the same transfer phenomena as found in previous studies employing these paradigms. The ordering of the transfer groups over 15 trials supports the results obtained by Twedt and Underwood (1959) in that the A-Br, A-C, and C-B transfer conditions all show negative transfer effects when compared to the C-D condition (t values of 6.09, 4.48, and 3.74, respectively, are significant at the .01 level). In addition, the A-Br group produced significantly greater negative transfer than the C-B group ($t = 2.35$, $p = .05$), whereas the A-C group did not ($t = .73$).

There is one notable difference between the transfer results obtained in the present study and those obtained in three earlier studies which employed the A-Br and the A-C paradigms (Besch & Reynolds, 1958; Porter & Duncan, 1953; Twedt & Underwood, 1959). In these three studies, the A-Br transfer paradigm resulted in significantly more negative transfer than did the A-C paradigm. In the present experiment, the A-Br and A-C groups do not differ significantly ($t = 1.61$) in the amount of negative transfer as measured by number of correct responses over 15 trials. Other results, however, are more in line with those obtained in the earlier studies: (*a*) Performance in the A-Br condition was below that of the A-C group for 13 of the 15 trials spent in second-list learning. (*b*) Intrusion data support earlier findings. If, during second-list learning, the subject responded to a second-list stimulus with the adjective that was *paired* with this stimulus in the first list, this error was classified as a first-list intrusion. Of all the errors made by the A-Br group during second-list learning, 28.84% were first-list intrusions. Only 2.09% of all A-C errors were such first-list intrusions. In terms of total number of overt errors over 15 trials,

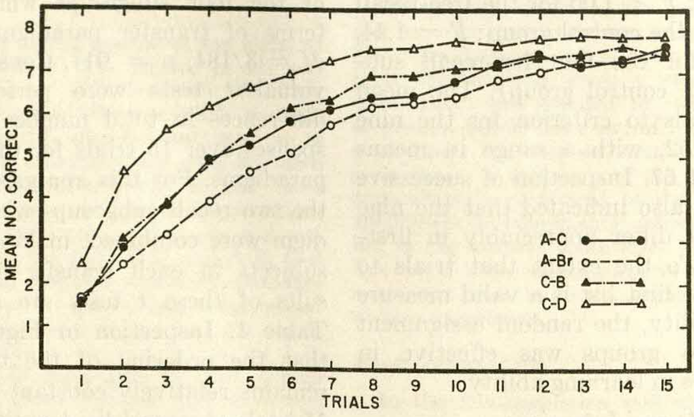


FIG. 1. Mean number of correct responses in second-list learning plotted over 15 trials as a function of transfer conditions. (Free- and list-recall subgroups of each paradigm have been combined.)

the A-Br group surpassed the A-C group with 1,016 as compared with 573 overt errors.

In general, then, the results concerning relative transfer indicate that the present study deals with essentially the same transfer phenomena as found in previous studies employing these four paradigms.

Methods Used in Scoring Recall of First-List Responses

Recall was scored in two different ways for each subject in the free-recall subgroups. The first method was the most stringent in that only those first-list adjectives paired with the appropriate stimulus were counted as correct. In the second, or liberal, scoring method, all first-list adjectives recalled by the subject were scored as correct regardless of placement. The stringent method described above was the only meaningful method for scoring recall in the list-recall subgroups since subjects in these subgroups were given a list of first-list adjectives to pair with first-list stimuli. The maximum possible recall score for any scoring method was 8.00.

The stringent and liberal methods of scoring recall in the free-recall subgroups and the stringent scoring method in the list-recall subgroups make three separate analyses of recall possible. Mean recall as determined in these three analyses is summarized in Table 2. Statistical evaluation appears in Table 3. The median test (Walker & Lev, 1953) has been used in these analyses because inspection of the recall scores revealed negatively skewed distributions for some of the groups involved in each analysis as well as heterogeneous variances. Where χ^2 values obtained in the median tests only approached significance, *t* tests were also performed on the data in an attempt to compensate for the lack of power associated with nonparametric statistical tests.

Free Recall: Stringent Scoring

The results provided by the stringent scoring method in the free-recall subgroups are summarized in Figure 2 and Tables 2 and 3. It appears that a loss in recall for the A-B list occurs for all free-recall transfer groups as compared to the control group. The median test gave a significant overall χ^2 of 35.62 ($p = .01$). Consequently, individual median tests were made comparing each transfer condition with the control group and with the other transfer conditions. The results of the individual median tests presented in Table 3 indicate that all transfer conditions produce a significant decrement in recall as compared to the control group when free recall is scored by the stringent method (the chi-square test was significant at the .05 level for the C-B condition and at the .01 level for each of the other conditions). Recall was significantly lower ($p = .05$ or better) in the A-C and A-Br groups than in the C-B and C-D conditions. Recall did not differ significantly between the A-C and A-Br conditions ($\chi^2 = .08$) nor in the C-B and C-D groups ($\chi^2 = 1.35$).

TABLE 2

MEAN NUMBER OF RESPONSES RECALLED FROM LIST 1 FOR EACH TRANSFER PARADIGM AS SCORED BY TWO METHODS[a]

Method	Paradigm				
	A-C	A-Br	C-B	C-D	Control
			Free-recall subgroup		
1	4.79	4.75	6.42	6.54	7.71
2	5.00	6.96	7.71	6.62	7.71
			List-recall subgroup		
1	6.54	5.58	6.87	7.83	—

[a] In Method 1 a response is scored as correct only if matched with the correct stimulus. In Method 2 a response is counted as correct if merely recalled.

In the introduction to the present study, it was hypothesized that two factors operated in the A-C condition to produce a

TABLE 3

CHI-SQUARE VALUES FOR INDIVIDUAL MEDIAN TESTS PERFORMED ON FREE-RECALL SUBGROUPS SCORED BY METHOD 1

Paradigm	A-Br	C-B	C-D	Control
A-C	.08	5.49*	5.37*	19.60**
A-Br	—	5.49*	6.86**	12.09**
C-B	—	—	1.35	5.37*
C-D	—	—	—	6.76**

* $p < .05$ (for $p < .05$ $\chi^2 = 3.8$, $df = 1$).
** $p < .01$ (for $p < .01$ $\chi^2 = 6.6$, $df = 1$).

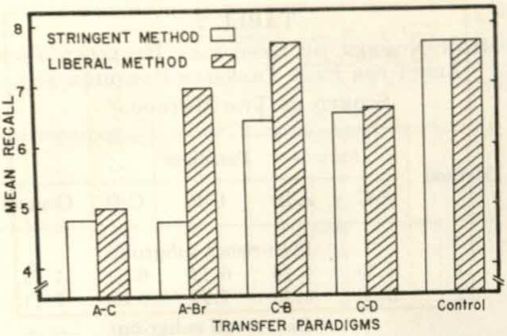


FIG. 2. Mean recall as measured by liberal and stringent scoring methods in the free-recall subgroups of the four transfer conditions.

decrement in immediate free recall: (*a*) extinction of first-list S-R associations; (*b*) extinction of associative connections presumably established between environmental stimuli and first-list responses. It was assumed that recall in the C-D condition would be hampered by the second of these factors, but not the first. The significant loss in recall in the C-D condition reflects, according to our theory, extinction of associative connections between contextual stimuli and first-list responses. It seems likely that such extinction also contributes to the retention loss exhibited in the A-C condition. Similarly, two factors were postulated as possible sources of RI in the A-Br condition: (*a*) extinction of first-list S-R associations and (*b*) extinction of first-list R-S associations. The significant decrement in recall obtained in the C-B condition appears to indicate the extinction of R-S associations. Again it seems reasonable to assume the operation of a similar extinctive process during second-list learning in the A-Br condition, since we have shown the implied relationship to exist for this condition.

The question which next arises is concerned with whether it is necessary to attribute part of the decrement in recall in the A-C and A-Br paradigms to extinction of first-list S-R associations or whether the factors indicated by analysis of the C-D and C-B conditions, respectively, are sufficient to account for the decrements. The results seem to support an interpretation in favor of the presence of the second, common factor (extinction of S-R associations) in the A-C and A-Br conditions. Loss in recall is significantly greater in the A-C than in the C-D condition and significantly greater in the A-Br condition than in the C-B condition. In addition, the increases in RI in the A-C and A-Br conditions as compared to the C-D and C-B conditions, respectively, are approximately equal as would be expected if S-R extinction were a common factor in the A-C and A-Br conditions and occurred to the same extent in both conditions. Further analyses discussed below also support the notion that there are two extinctionlike factors operating during second-list learning in both the A-C and A-Br paradigms to produce a decrement in recall. So far, our results appear to indicate that decrement in recall is directly related to the number of extinction factors operative in second-list learning. RI is significantly greater in the two paradigms for which two extinction factors have been postulated than in the two paradigms for which only one such factor has been postulated. The remaining two analyses discussed below support the results obtained with stringent scoring in the free-recall subgroups.

List Recall: Stringent Scoring

The recall scores of the list-recall groups should supply information concerning the weakening of first-list associative connections independent of any extinction effects in the response-recall phase. Subjects in the list-recall subgroups were given the first-list adjectives and were instructed to pair them with the appropriate first-list stimuli. This procedure assures perfect response-recall in all groups and should eliminate the negative effects of having to learn and retain two different sets of responses in the A-C and C-D conditions.

Comparison of recall as measured by the stringent method in the free-recall and list-recall subgroups seems to support the conclusions reached in the previous section concerning extinction of contextual associations in the A-C and C-D conditions. Inspection of the mean recall scores presented in Table 2 indicates that supplying the subjects with first-list reponses increased mean recall in all transfer groups. Median tests indicate that this increase was significant, however, only in the case of the C-D group ($\chi^2 = 17.42$, $df = 1$, $p = .01$). The increase approaches significance in the A-C condition where a χ^2 value of 3.08 was obtained, a value of 3.80 being required for significance at the .05 level. A *t* test performed on the A-C subgroups indicated a significant increase in recall when first-list responses are supplied ($t = 2.98$, $df = 46$, $p = .01$). It would appear, then, that learning two different sets of responses in the A-C and C-D conditions has a detrimental effect on recall and that this effect is eliminated in the list-recall condition when the subjects are supplied with the set of responses learned in the first list.

This interpretation must be viewed with caution because these increases in recall may be due in part, at least, to increased opportunities for guessing when first-list responses are supplied. No method of assessing the role played by guessing in these two groups has been provided for in this experiment. On the other hand, the results which were discussed in the preceding section on free-recall and those which will be discussed in the following section may be interpreted as supporting the extinction of associations to contextual cues. Thus, while the significant increase in recall in the A-C and C-D conditions is not conclusive in the present analysis, this increase is consistent with other analyses which point more conclusively toward extinction in the response-recall phase of paired-associate learning. The increases in recall in the A-Br and C-B conditions are slight and may be attributed statistically to random error.

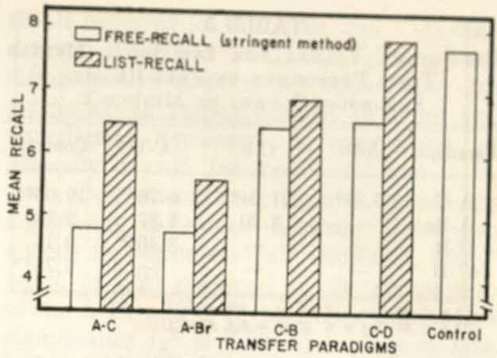


FIG. 3. Mean recall as stringently scored in the free- and list-recall subgroups of the four transfer paradigms.

Turning next to list-recall itself, the results of an analysis of that measure are summarized in Figure 3 and Tables 2 and 4. The overall median test produced a significant χ^2 of 21.92 ($p = .01$). The results of the individual median tests are presented in Table 4. Before turning to these individual comparisons, it should be noted that, with response-recall effects eliminated, any decrement in recall should be the result of the weakening or extinction of first-list S-R and/or R-S associative connections. The C-D group may serve as the control for determining whether or not such extinction has occurred in the three other transfer paradigms. Theoretically, there is no opportunity for the disruption of first-list forward and backward associations in the C-D paradigm, hence, mean recall for the C-D subgroup in the present analysis should be perfect, i.e., 8.00. Inspection of the recall scores in the C-D subgroup indicates that 22 subjects had recall scores of 8, and 2 subjects had recall scores of 6, producing a mean recall of 7.83. Hence, theory would appear to be supported, and the C-D subgroup may be used as a control for the other three transfer groups.

With response-recall assured, there should be only one negative factor operating in the A-C condition, the weakening of S-R associations, an effect which may be assessed by a comparison of the performance of the A-C group with that of the C-D group. The significant difference in this comparison ($\chi^2 = 8.54$, $df = 1$)

TABLE 4

CHI-SQUARE VALUES FOR INDIVIDUAL MEDIAN TESTS PERFORMED ON LIST-RECALL SUBGROUPS SCORED BY METHOD 1

Paradigm	A-Br	C-B	C-D
A-C	3.09	.09	8.54**
A-Br	—	5.37*	21.94**
C-B	—	—	7.11**

* $p < .05$ (for $p < .05$ $\chi^2 = 3.8$, $df = 1$).
** $p < .01$ (for $p < .01$ $\chi^2 = 6.6$, $df = 1$).

TABLE 5

CHI-SQUARE VALUES FOR INDIVIDUAL MEDIAN TESTS PERFORMED ON FREE-RECALL SUBGROUPS SCORED BY METHOD 2

Paradigm	A-Br	C-B	C-D	Control
A-C	5.34*	21.94**	6.76**	19.60**
A-Br	—	3.20	1.37	2.18
C-B	—	—	8.40**	0.11
C-D	—	—	—	6.76**

* $p < .05$ (for $p < .05$ $\chi^2 = 3.8$, $df = 1$).
** $p < .01$ (for $p < .01$ $\chi^2 = 6.6$, $df = 1$).

indicates the presence of this factor and verifies the interpretation presented in the preceding section. The significant chi-square value ($\chi^2 = 7.11$, $df = 1$) obtained for the C-B condition indicates the extinction of R-S associations in the C-B condition and again supports the findings obtained with the stringent scoring method in the free-recall C-B subgroup. A comparison of recall in the A-C and C-B conditions indicates that the extinction of S-R and R-S associations produces approximately the same degree of RI (χ^2 comparing A-C and C-B conditions has a value of .09). Both of the negative factors involving first-list associative connections were assumed to operate in the A-Br condition: the extinction of S-R and the extinction of R-S associative connections. The significant χ^2 of 21.94 obtained for the A-Br and C-D conditions indicates that extinction of first-list associative connections has occurred during second-list learning in the A-Br condition. If the A-C condition as scored by the stringent scoring method in list-recall may be assumed to be an estimate of S-R extinction in the A-Br group independent of R-S extinction, S-R extinction has been verified for the A-Br condition. The fact that recall is lower in the A-Br group than in the A-C ($\chi^2 = 3.09$ is not significant; t test comparing A-C and A-Br was also not significant) and C-B conditions ($\chi^2 = 5.37$, $df = 1$, $p = .05$) each of which have only one extinction factor operating in list-recall, supports the hypothesis that two extinctionlike processes operate in the A-Br condition. In fact, the decrement in recall attributable to the extinction of S-R associative connections as estimated in the A-C condition and the decrement in recall attributable to the extinction of R-S associative connections as estimated by the C-B condition, when combined in an additive fashion, are sufficient to account for the decrement observed in the A-Br condition.

As was found with the stringent scoring method in the free-recall subgroups, decrement in recall measured by the stringent scoring method in list-recall appears to be directly related to the number of extinction factors assumed to be operative during second-list learning. Results obtained with the stringent scoring method in the list-recall subgroups indicated that this general finding is true when response-recall factors are eliminated, and only the extinction of associative connections between first-list stimuli and responses is involved.

Free Recall: Liberal Scoring

The liberal method of scoring in the free-recall subgroups provides further information concerning extinction in the response-recall phase in the A-C and C-D conditions. In this method of scoring, a response is counted as correct if it is merely recalled. The overall median test produced a significant χ^2 of 29.27 ($p = .01$). Individual median tests are summarized in Table 5. Inspection of mean recall in Figure 2 and Table 2 indicates a loss in recall for all transfer conditions except the C-B condition when comparisons are made with the control group. The individual median tests summarized in Table 5 indicate that this decrement in response-recall is significant only for the A-C and C-D conditions. Thus, a general loss in response-recall is significant only for those two paradigms for which it was hypothesized as playing an important role. In addition, response-recall in the A-C and C-D paradigms was significantly lower than in the A-Br and C-B conditions (see Table 5) for which no response-recall extinction was hypothesized. These findings support the results obtained by comparing free-recall as scored by the stringent method with list-recall in the preceding section.

At first glance, it might appear that the extinction produced by the learning of a new set of responses in the second list occurs to a greater extent in the A-C than in the C-D condition. A significant χ^2 value of 19.60 ($p = .01$) indicates that response-recall is lower in the A-C than in the C-D condition. This does not mean, however, that greater extinction effects in the response-recall phase are produced in the A-C condition. Although the associative connections between environmental stimuli and first-list responses are weakened in the C-D condition, first-list S-R associative connections presumably have been unaffected. Thus, recall of a given first-list response may be mediated by its first-list S-R associative connection even though its associative connections with environmental stimuli have been weakened. In the A-C condition, on the other hand, evidence summarized in previous analyses indicates that S-R associative connections have been weakened as well. Thus, in the A-C condition, a given response may have lost both its associative connection with environmental stimuli and with its first-list stimulus and, without either of these two connections, cannot be recalled. Hence, the significant difference in recall in the A-C and C-D conditions when scored by the liberal method is more likely due to the weakening of S-R associations in the A-C condition than to greater extinction effects in the response-recall phase of the A-C condition.

An inspection of Table 2 shows that for the A-C and C-D free-recall groups there is little difference in mean recall produced by the two scoring methods. Similar results were obtained for the A-C condition in the Barnes-Underwood (1959) study. This result was interpreted in that study as indicating that if a subject recalled a response, he could also identify it correctly as to stimulus. This led to the speculation presented in the introduction that perhaps the subject could correctly pair *all* first-list responses with first-list stimuli in recall if first-list responses were supplied. Results obtained with the A-C condition in list-recall and summarized in the preceding section indicated that this was not the case.

Comparison of recall in the A-Br and C-B conditions with the control group indicated that no significant loss in response-recall occurred in the A-Br and C-B conditions. This was as it should be since, theoretically, there is no opportunity for extinction in response-recall in these two paradigms. Similarly, a comparison of liberally scored free-recall for the A-Br and C-B conditions should reveal no significant difference between these two groups in response-recall. However, the χ^2 value of 3.20 obtained in the comparison of the A-Br and C-B conditions approaches significance ($\chi^2 = 3.80$, $df = 1$, $p = .05$). A t test (the t test appropriate for unequal variances with special formula for determining degrees of freedom was used—Walker & Lev, 1953, p. 155) was performed on the differences in mean recall for these two conditions and indicated a significant difference in response-recall ($t = 2.40$, $df = 30$, $p < .05$). There is no ready explanation for this difference. It may be that the extinction of both S-R and R-S associative connections in the A-Br condition confused the subjects to such an extent that they hesitated to guess even though instructed to do so.

Free Association

The free-association task was designed to determine whether the "extinguished" first-list responses were lost from the subjects' repertoires merely in terms of the specific associations which were developed in first-list learning and then extinguished during second-list learning or whether a more general extinction effect was produced so that they could not be used as responses to other stimuli where they would be appropriate. The "other" stimuli used in the present experiment were eight common adjectives for which first-list responses were appropriate synonyms. The results of the present study indicate that the operations employed generally *facilitated* the use of first-list responses in the free-association task for all subgroups. First-list responses were used by 50.0%–98.5% of the subjects in the different groups in contrast to the range of 30.0%–50.0% obtained with pilot subjects.

DISCUSSION

Theoretically, three factors were identified in the introduction as involved in the production of RI in the immediate written free recall of four transfer paradigms: the A-B, A-C; the A-B, A-Br; the A-B, C-B; and the A-B, C-D paradigms. The results of the present study have been interpreted only in terms of their implications for the existence of the three negative factors hypothesized as responsible for RI in immediate recall. This does not mean that the author wishes to imply that these are the only factors which are involved in RI or that they necessarily constitute the most parsimonious explanation of the results. The results, however, may be interpreted as favorable evidence for the operation of such extinctionlike factors. The evidence indicates that when an A-B, A-C relationship exists between sets of associative connections, an extinction of the first set of associative connections learned occurs as a consequence of having learned the second set.

One or more of these three processes was identified as a possible source of RI in written free recall for each of the four transfer paradigms studied in the present experiment. Consequently, RI was predicted for free recall as scored by the stringent method in each of the four transfer paradigms. This prediction was verified. In addition, recall was significantly lower in those groups which theoretically had two negative factors operating during second-list learning (the A-B, A-C and A-B, A-Br paradigms) than in those groups for which only one negative factor was indicated (the A-B, C-B and A-B, C-D paradigms). Specific evidence for each of the three sources of RI in immediate recall is summarized separately for each transfer paradigm.

The A-B, C-D Paradigm

Almost perfect recall (mean recall of 7.83 versus perfect recall of 8.00) was obtained in the C-D condition when the subjects were supplied with the set of responses learned in the first list. In other words, when the subject had first-list responses at his disposal in recall, he could pair them properly with first-list stimuli. This indicates that the extinction of first-list S-R and R-S associations does not play an important role in recall in the A-B, C-D paradigm. On the other hand, the significant decrement in recall found for the C-D condition in free-recall scored in either of two ways indicates that the learning of two different sets of responses produces extinction in the response-recall phase of paired-associate learning. Thus, after learning a new set of responses in a second list, the subject has difficulty in recalling the responses which made up the set learned in the first list. However, the subject can pair first-list responses correctly with first-list stimuli when responses are supplied by the experimenter.

The RI found in the recall scores of the A-B, C-D paradigm appears to indicate that: (*a*) associative connections are formed between a sample of environmental or contextual cues and list responses; (*b*) extinction of these connections occurs as a consequence of the formation of similar associative connections with a new set of responses during second-list learning; (*c*) these associative connections are important in mediating response-recall, per se, independent of any associative connections formed between list stimuli and responses.

To the extent that recall in the A-B, C-D condition may be assumed to serve as an independent estimate of the effect of learning two separate response sets in the A-B, A-C paradigm, extinction in the response-recall phase of paired-associate learning has been demonstrated for the A-B, A-C paradigm.

The A-B, A-C Paradigm

Two negative factors in recall were identified theoretically for the A-B, A-C paradigm: (*a*) extinction of first-list S-R associations and (*b*) extinction of associative connections established between environmental stimuli and first-list responses, i.e., extinction in the response-recall phase. The second factor has been verified independently of the extinction of S-R associative connections in the A-B, C-D paradigm as discussed in the preceding section. The fact that there was a fairly large increase (a significant increase as measured by a *t* test; a nonsignificant increase as measured by the median test) in recall for the A-C condition when the subjects were supplied with first-list responses in the list-recall subgroups as compared to stringently scored recall in the free-recall subgroups is a further indication that losses in response-recall, per se, are involved to a certain extent in the A-B, A-C paradigm. Liberal scoring of free-recall also indicated a significant loss in response-recall for the A-C condition.

Evidence for the extinction of first-list S-R associations was found in recall as stringently scored in the free-recall and list-recall subgroups of the A-C condition. Free-recall stringently scored in the A-C condition was significantly below that obtained in the C-D condition. Assuming free-recall in the C-D condition accurately estimated the effect of extinction in the response-recall phase for the A-C condition, this significant difference may be taken as an indication of the presence of a second negative factor operating in the A-C condition over and above extinction in response-recall. This second factor was identified theoretically as the extinction of first-list S-R associations. An independent verification of first-list S-R extinction was obtained in the list-recall subgroup of the A-C condition. It will be remembered that the effects of extinction in the response-recall phase were eliminated in the list-recall subgroup by supplying the subjects with the set of first-list responses. Recall in this A-C subgroup was significantly lower than in the appropriate control condition indicating that extinction of associative connections between first-list stimuli and responses is involved in the A-B, A-C paradigm. Since the R-S or backward associative connections involved in this paradigm presumably have zero transfer value, it may be inferred that the factor responsible for the loss in recall in the list-recall subgroup is the extinction of first-list S-R or forward associative connections.

Since an A-B, A-C relationship exists among the first- and second-list S-R or forward associations in the A-B, A-Br paradigm, the significant loss in recall in the list-recall subgroup of the A-C condition may be used as an independent estimate of the extinction of S-R associations in the A-B, A-Br paradigm.

Results obtained in the A-B, A-C paradigm appear to support the conclusion arrived at in the Barnes-Underwood study that extinction of first-list S-R associations is responsible for much of the decrement found in immediate free-recall for this paradigm. They also support results obtained in the Barnes-Underwood study in still another way. It was noted in the earlier study that the subjects did not confuse first- and second-list responses in written recall in the A-B, A-C paradigm. If the subject recalled a response, he could identify it correctly as to list. Similar results were obtained in the free-recall subgroups of the A-C and C-D conditions. Subjects used a second-list response in recall only four times in the A-C condition and only twice in the C-D condition. This indicates that confusion of response systems plays little part in the production of RI in immediate free recall for the A-C and C-D paradigms.

The A-B, C-B Paradigm

A significant decrement in recall was obtained in the C-B condition as compared with the appropriate control in both free-recall and list-recall as scored by the stringent scoring method, indicating a weakening of first-list associative connections. Liberal scoring of the free-recall subgroup of the C-B condition indicates that extinction in the response-recall phase, per se, is not a factor in the A-B, C-B paradigm as might be expected from the nature of the paradigm. In addition, supplying subjects in this condition with first-list responses does not significantly increase recall. Thus, it may be concluded that extinction of associative connections established between first-list stimuli and first-list responses alone, is responsible for recall losses in the A-B, C-B paradigm.

Since S-R or forward associations are assumed to have zero transfer value due to the materials used in the present experiment, the presence of RI in the A-B, C-B paradigm may be attributed to the extinction or weakening of first-list R-S associations. In other words, the RI obtained in the A-B, C-B paradigm can be taken to mean all of the following: (*a*) backward or R-S associations were formed in each of the two lists learned; (*b*) extinction of List 1 backward associations occurred as a result of second-list learning; (*c*) the existence of an A-B, A-C relationship between R-S associations of two lists is an extinctive one; (*d*) backward or R-S associations are important in recall.

At this point, a question may be raised as to how backward associations might be important in recall. Two possibilities suggest themselves: First, backward associations may serve as a means of "checking" responses recalled via first-list S-R associations. For example, in the C-B condition, a given first-list stimulus may elicit, via first-list S-R association, the response that was paired with it in first-list learning. The R-S or backward association formed during first-list learning may act as a check for any possible confusion of first-list stimuli and their appropriate first-list responses. Extinction of first-list R-S associations in the C-B condition removes this possible source of checking in recall. Second, first-list backward associations may mediate recall when first-list S-R associations have been weakened or extinguished. For example, in the A-C condition, a first-list stimulus and response might be correctly paired because the subject has available the R-S association which was formed between that response and that stimulus in first-list learning even though the forward or S-R association has been weakened or extinguished in second-list learning. Thus, backward associations may serve either or both of two functions in recall: (*a*) they may serve as a check on possible confusion in recall mediated by first-list S-R associations; (*b*) they may mediate the pairing of first-list stimuli and responses when first-list S-R associations have been extinguished or weakened.

It was further assumed that the presence of RI in the A-B, C-B paradigm might be used as an independent verification of the extinction of R-S associations formed in the A-B, A-Br paradigm. To the extent that this assumption is a valid one, the extinction of R-S associations has been verified for the A-B, A-Br paradigm.

The A-B, A-Br Paradigm

Two factors were postulated as producing recall decrement in the A-B, A-Br paradigm: (*a*) extinction of first-list S-R or forward associative connections and (*b*) extinction of first-list R-S or backward associative connections. This first factor has been verified independently in the list-recall subgroup of the A-C condition. The existence of the second factor has been supported by the free- and list-recall results in the C-B condition. These independent verifications of the two processes have been discussed above.

The presence of these two factors is also indicated by recall in the A-Br condition itself. If free-recall in the C-B condition may be considered an accurate estimate of the extinction of first-list R-S associations in the A-B, A-Br paradigm, then the significant difference found between the A-Br and C-B conditions in free-recall as scored by the stringent method indicates the presence of a second negative factor in A-Br recall. Theoretically, this second factor can only be extinction of first-list S-R associative connections. Since the effects of extinction in response-recall have been eliminated in list-recall, the A-C and C-B conditions each have only one negative factor operating in recall, extinction of first-list S-R and R-S associative connections, respectively. The A-Br condition, on the other hand, presumably has both factors operating to reduce list-recall scores. A comparison of list-recall subgroups indicated that recall in the A-Br condition was lower than recall in the A-C condition and significantly lower than recall in the C-B condition. In addition, the magnitude of the decrement found in list-recall of the A-Br condition is approximately equivalent to the sum of the magnitudes of the decrements found in the A-C and C-B conditions. It would appear, then, that as far as first-list associative factors are concerned, RI in immediate recall in the A-B, A-Br paradigm can be accounted for in terms of the extinction of first-list S-R and R-S associative connections.

As might be expected from the nature of the paradigm, extinction in the response-recall phase of paired-associate learning is not a factor in recall in the A-B, A-Br paradigm. Supplying subjects in this condition with first-list responses did not significantly raise recall. Liberal scoring in the free-recall subgroup also revealed no significant loss in response-recall.

Conclusion

The results of the present study would seem to indicate that paired-associate recall may be logically and empirically divided into two phases or stages comparable to the two stages identified in paired-associate learning (Underwood et al., 1959), namely, a response-recall phase and an association phase. Evidence has also been obtained which indicates that a loss in recall may occur in one phase independent of loss in the other phase. Thus, a significant loss occurred in the response-recall phase of the A-B, C-D paradigm yet no loss in recall was found in the associative phase when loss in the response-recall phase was eliminated. Similarly, a significant decrement in the associative phase of recall was determined for the A-B, C-B and A-B, A-Br paradigms, whereas the response-recall phase remained unaffected. In addition, it would appear that loss in recall in the associative phase may be produced by either of two extinctionlike processes or a combination of these two processes: a weakening or extinction of first-list S-R associative connections and extinction of first-list R-S associative connections. In order to account for losses in response-recall, an extinction or weakening of associative connections which were formed between environmental stimuli and first-list responses during first-list learning was hypothesized. Assuming these explanations of recall decrement are valid ones, the results also seem to indicate that the presence of an A-B, A-C relationship between two sets of associative connections is an "extinctive" one, regardless of the type of association. That is, an A-B, A-C relationship appears to produce a loss in recall when the associations involved are S-R or forward associations, R-S or backward associations, and when the associations involved are between a sample of environmental cues and list responses. Thus, the results of the present experiment support those obtained in the Barnes-Underwood study for the A-B, A-C paradigm and confirms the conclusions concerning extinction of first-list responses. The present study also indicates that this extinction can be separated into two phases for the A-B, A-C paradigm and extends the findings to three other transfer paradigms.

SUMMARY

This study was concerned with the "fate" of first-list associations as determined by immediate written free recall in four transfer paradigms: A-B, A-C (stimuli identical; responses different); A-B, A-Br (stimuli and responses identical, but re-paired in the second list); A-B, C-B (stimuli different; responses identical); A-B, C-D (stimuli and responses both different). The lists consisted of eight paired-associates with nonsense syllables as stimuli and adjectives as responses. The first list was learned to a criterion of one perfect trial. A 1-minute rest interval separated first- and second-list learning. Subjects were stopped after 15 anticipation trials on List 2 and were given immediate written free recall in which they were instructed to write first-list responses after the appropriate first-list stimuli which were presented on a mimeographed recall sheet. Subjects were allowed 2 minutes for written recall. Half of the subjects in each transfer condition were given a list of the responses learned in List 1 and were instructed to pair them with first-list stimuli. These subjects formed the list-recall subgroups of the four transfer paradigms. The remaining subjects in each transfer condition formed the free-recall subgroups.

The two conditions of recall in the four transfer paradigms form eight independent groups of 24 subjects each. In addition, a control group of 24 subjects was run as a check on normal forgetting. All subjects received a "free-association" task follow-

ing written recall in an attempt to determine the extent or generality of the loss of first-list responses from the subject's repertoire produced by factors operating in the transfer paradigms during second-list learning.

Theoretically, three factors were identified as involved in the production of RI in immediate written free recall of the four transfer paradigms: (*a*) extinction of first-list S-R or forward associative connections; (*b*) extinction of first-list R-S or backward associations; and (*c*) extinction in response-recall, i.e., extinction of associative connections formed between environmental or contextual stimuli and first-list responses. The first and third factors were hypothesized as operating in the A-B, A-C paradigm. The first and second factors were hypothesized as operating in the A-B, A-Br paradigm. The A-B, C-B and A-B, C-D paradigms were included for independent verification of Factors *b* and *c*, respectively. Any one or a combination of these three factors might produce decrements in free-recall. Hence, RI was predicted for the free-recall subgroups of all four transfer paradigms. In list-recall, however, the effect of the third factor is presumably eliminated. Thus, RI was predicted in list-recall for all but the A-B, C-D condition. These predictions were verified. In addition, results verified all the hypotheses concerning extinctionlike factors operating during second-list learning in the four transfer paradigms.

The results would seem to indicate that paired-associate recall may be divided into two stages: a response-recall phase and an association phase. A loss in recall apparently may occur in one phase independent of loss in the other phase. Loss in recall in the associative phase may be produced by Factors *a* and/or *b* listed above. Losses in response-recall were attributed to factor *c*.

Results in the free-association task indicated a facilitation rather than a generalized extinction of the responses used in the paired-associate learning task.

REFERENCES

BARNES, JEAN M. "Fate" revisited. Unpublished doctoral dissertation, Northwestern University, 1960.

BARNES, JEAN M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

BESCH, N. F., & REYNOLDS, W. F. Associative interference in verbal paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 554–558.

BILODEAU, I. McD., & SCHLOSBERG, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *J. exp. Psychol.*, 1951, **41**, 199–204.

BUGELSKI, B. R., & CADWALLADER, T. C. A reappraisal of the transfer and retroaction surface. *J. exp. Psychol.*, 1956, **52**, 360–366.

GREENSPOON, J., & RANYARD, R. Stimulus conditions and retroactive inhibition. *J. exp. Psychol.*, 1957, **53**, 55–59.

HAAGEN, C. H. Synonymity, vividness, familiarity, and associative value ratings of 400 pairs of common adjectives. *J. exp. Psychol.*, 1949, **27**, 453–463.

HAMILTON, R. J. Retroactive facilitation as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1943, **32**, 363–376.

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

JENKINS, W. O., & POSTMAN, L. An experimental analysis of set in rote learning: Retroactive inhibition as a function of changing set. *J. exp. Psychol.*, 1949, **39**, 69–72.

LESTER, O. P. Mental set in relation to retroactive inhibition. *J. exp. Psychol.*, 1932, **15**, 681–699.

MANDLER, G. Response factors in human learning. *Psych. Rev.*, 1954, **61**, 235–244.

MANDLER, G., & HEINEMAN, S. H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **52**, 39–46.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Press, 1953.

PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

POSTMAN, L., & POSTMAN, D. L. Change in set as a determinant of retroactive inhibition. *Amer. J. Psychol.*, 1948, **61**, 236–242.

ROTHKOPF, E. A deduction from an extinction-inhibition account of retroactive inhibition. *J. exp. Psychol.*, 1957, **53**, 207–213.

SALTZ, E. A single theory for reminiscence, act regression, and other phenomena. *Psychol. Rev.*, 1953, **60**, 159–171. (a)

SALTZ, E. Act regression as a special case of retroactive inhibition and functionally related to stimulus generalization. *J. exp. Psychol.*, 1953, **45**, 394–400. (b)

STEVENS, S. S. (Ed.) *Handbook of experimental psychology*. New York: Wiley, 1951.

THUNE, L. E. Reproductive interference following "appropriate" and "inappropriate" warm-up activities. *J. exp. Psychol.*, 1958, **55**, 535–542.

TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

WATERS, R. H., & PEEL, Z. E. Similarity in the form of original and interpolated learning and retroactive inhibition. *Amer. J. Psychol.*, 1935, **47**, 477–481.

WHITELY, P. L. The dependence of learning and recall upon prior intellectual activities. *J. exp. Psychol.*, 1927, **10**, 489–508.

YOUNG, R. K., & UNDERWOOD, B. J. Transfer in verbal materials with dissimilar stimuli and response similarity varied. *J. exp. Psychol.*, 1954, **47**, 153–159.

(Received January 10, 1964)

BUREAU OF EDUCATIONAL AND PSYCHOLOGICAL RESEARCH
GOVT. OF W. BENGAL
LIBRARY

ABILITATION·I [illegible] IA
CALC[illegible]A